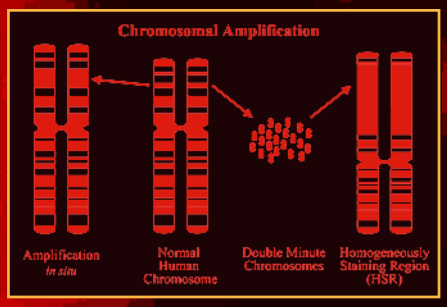
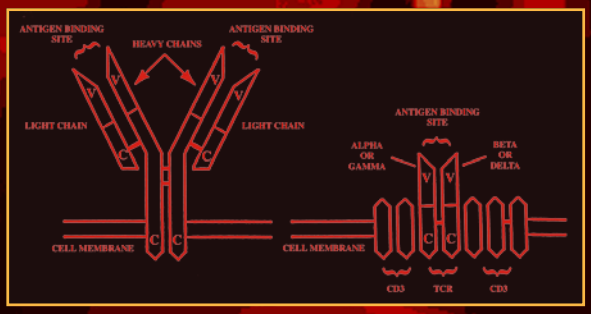
Molecular Diagnostics

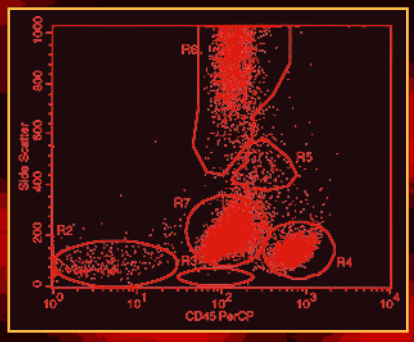
For the Clinical Laboratorian

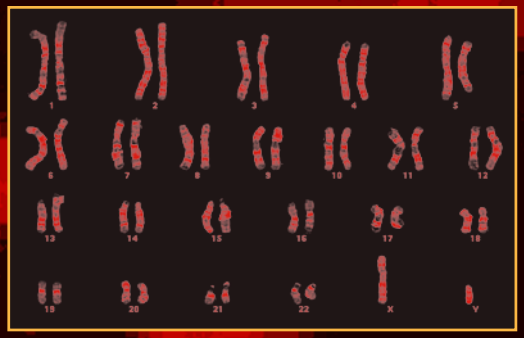
SECOND EDITION

Edited by
William B. Coleman
Gregory J. Tsongalis









HUMANA PRESS

Includes eBook/PDA on CD-ROM

Molecular Diagnostics

Molecular Diagnostics

For the Clinical Laboratorian

Second Edition

Edited by

William B. Coleman

Department of Pathology and Laboratory Medicine
University of North Carolina School of Medicine
Chapel Hill, NC

and

Gregory J. Tsongalis

Department of Pathology
Dartmouth Medical School
Dartmouth-Hitchcock Medical Center
Lebanon, NH

Foreword by

Lawrence M. Silverman

Department of Pathology and Laboratory Medicine
University of Virginia
Charlottesville, VA

Humana Press
Totowa, New Jersey

© 2006 Humana Press Inc.
999 Riverview Drive, Suite 208
Totowa, New Jersey 07512

humanapress.com

For additional copies, pricing for bulk purchases, and/or information about other Humana titles, contact Humana at the above address or at any of the following numbers: Tel.: 973-256-1699; Fax: 973-256-8341; E-mail:orders@humanapr.com; Website: http://humanapress.com

All rights reserved.

No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise without written permission from the Publisher.

All articles, comments, opinions, conclusions, or recommendations are those of the author(s), and do not necessarily reflect the views of the publisher.

Due diligence has been taken by the publishers, editors, and authors of this book to assure the accuracy of the information published and to describe generally accepted practices. The contributors herein have carefully checked to ensure that the drug selections and dosages set forth in this text are accurate and in accord with the standards accepted at the time of publication. Notwithstanding, as new research, changes in government regulations, and knowledge from clinical experience relating to drug therapy and drug reactions constantly occurs, the reader is advised to check the product information provided by the manufacturer of each drug for any change in dosages or for additional warnings and contraindications. This is of utmost importance when the recommended drug herein is a new or infrequently used drug. It is the responsibility of the treating physician to determine dosages and treatment strategies for individual patients. Further it is the responsibility of the health care provider to ascertain the Food and Drug Administration status of each drug or device used in their clinical practice. The publisher, editors, and authors are not responsible for errors or omissions or for any consequences from the application of the information presented in this book and make no warranty, express or implied, with respect to the contents in this publication.

Cover illustrations: *Background*—Figure 4 from Chapter 15. *Foreground*—Figure 6B from Chapter 2; Figure 5B from Chapter 13; Figure 1 from Chapter 22; and Figure 3A,B from Chapter 32.

Cover design by Patricia F. Cleary

This publication is printed on acid-free paper. ∞
ANSI Z39.48-1984 (American National Standards Institute) Permanence of Paper for Printed Library Materials.

Photocopy Authorization Policy:
Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by Humana Press Inc., provided that the base fee of US $30.00 per copy is paid directly to the Copyright Clearance Center at 222 Rosewood Drive, Danvers, MA 01923. For those organizations that have been granted a photocopy license from the CCC, a separate system of payment has been arranged and is acceptable to Humana Press Inc. The fee code for users of the Transactional Reporting Service is: [1-58829-356-4/06 $30.00].

Printed in the United States of America. 10 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging-in-Publication Data
Molecular diagnostics : for the clinical laboratorian / edited by William B. Coleman and Gregory J. Tsongalis.— 2nd ed.
p. cm.
Previously published in 1997.
Includes bibliographical references and index.
ISBN 1-58829-356-4 (alk. paper) eISBN 1-59259-928-1
1. Molecular diagnosis. 2. Pathology, Molecular. I. Coleman, William B.
II. Tsongalis, Gregory J.
RB43.7.M6338 2004
616.07'56—dc22 2004023555

Dedication

First and foremost, we would like to dedicate this book to our wives, Monty and Nancy, for their unprecedented support of our careers and families.

To all of our colleagues, students, trainees, and mentors over the many years, we thank you for your continued support.

FOREWORD

In 1994 I coedited a book, *Molecular Pathology: Approaches to Diagnosing Human Disease in the Clinical Laboratory*, that coined the phrase "molecular pathology" to refer to applications of molecular biology in the traditional areas of laboratory medicine. That compilation of clinical molecular techniques included 11 chapters and an epilogue on "New Directions for the Clinical Laboratory." Chapter headings included: automation, neoplasia, heritable diseases, and infectious diseases.

Three years later, in 1997, Bill Coleman and Greg Tsongalis edited *Molecular Diagnostics: For the Clinical Laboratorian*, which contained 18 chapters and introduced the new areas of RT-PCR and *In Situ* PCR, and included a section on gene therapy. Now, ten years after the introduction of molecular pathology, we are looking at the second edition of Bill Coleman and Greg Tsongalis' text.

This second edition now comprises 44 chapters, and contains new chapters on bioinformatics, microarrays, methylation assays, FISH, laser capture microdissection, quality assurance, chimerism studies for bone marrow transplantation, and separate chapters on genetic counseling and ELSI (ethical, legal, and social issues). These new chapters represent developments in the past decade that have fundamentally changed the scope from molecular pathology to molecular diagnostics, highlighting the changing role of the clinical laboratorians who direct these efforts. It should be evident that molecular pathology and diagnostics impact almost every conceivable subspecialty in laboratory medicine and, particularly in the case of pharmacogenomics, have led to the development of new areas of investigation.

What is less evident is the burden placed upon the laboratorians directing these efforts. Issues regarding training, certification, continuing education, and reimbursement (just to name a few) have demanded enormous time and effort from professional organizations and governmental agencies. In addition, because conventional approaches cannot always be applied to molecular diagnostics, unique solutions for quality assurance and quality control must be developed. These challenges have engendered committees, subcommittees, taskforces, and workgroups from regulatory agencies and professional organizations, representing worldwide constituencies. As in any situation involving so many players, consensus-building and communication are mandatory. To these ends, professional journals and textbooks are our best hope for remaining current with this rapidly changing field. One needs only to compare the wealth of knowledge in this current edition with our efforts of only ten years ago to appreciate the magnitude of this challenge.

Lawrence M. Silverman, PhD

PREFACE

It has been almost ten years since the concept for producing the first edition of *Molecular Diagnostics: For the Clinical Laboratorian* was conceived. In those ten years the field of molecular pathology and diagnostics has exploded as many predicted. The clinical diagnostic laboratory continues to function as the playing field for this expansion that includes vast and dynamic changes in test menus, instrumentation, and clinical applications. The impact of this field on the routine practice of clinical medicine and management of patients continues to be felt as new developments that span all areas of laboratory medicine exceed our expectations.

The success of this technology in a clinical setting is highly dependent upon the training of well-qualified technologists, residents, and clinicians alike, who will not only have to perform and interpret results of these tests but also understand the limitations of the technology and resulting clinical implications. The production of this second edition is a testament to our passion and commitment for the teaching and training of qualified individuals who wish to embark on this journey. The numerous training programs, educational venues, and board certification examinations that have evolved over the past ten years also sends a strong vow of commitment by others in the field to ensure the successful use of these new tools in supporting the best possible patient care that is available.

The second edition of *Molecular Diagnostics: For the Clinical Laboratorian* begins with a historical perspective of laboratory medicine followed by an overview of basic molecular biology techniques and concepts. Part III provides a more in depth examination of some advanced molecular technologies and their potential uses. Part IV describes other technologies found in the clinical laboratory that can complement or be complemented by molecular diagnostic technologies. The increasing need for awareness and practice of quality assurance in this field led us to include a complete section (Part V) that examines some of these issues. Although the first edition included clinical applications all in one section, the increased number of applications led us to develop separate sections for genetic disease, human cancers, infectious diseases, and identity testing (Parts VI–IX). Finally, the book concludes with a section on genetic counseling and ethical/social issues involved with nucleic acid testing.

Although no such book could possibly be all encompassing in such a rapidly developing field, we feel that the material covered herein will provide the reader with an excellent overview.

William B. Coleman
Gregory J. Tsongalis

CONTENTS

Color Plates

Color Plate 1: *Fig. 1 (Ch. 11, p. 124). cDNA microarray manufacturing process.*

Color Plate 2: *Fig. 2 (Ch. 11, p. 125). Example of cDNA microarray image file.*

Color Plate 3: *Fig. 11 (Ch. 11, p. 138). Dendrogram.*

Color Plate 4: *Fig. 1 (Ch. 15, p. 191).* In situ *hybridization for detection of mRNA.*

Color Plate 5: *Fig. 2 (Ch. 15, p. 194). Different types of DNA FISH probes.*

Color Plate 6: *Fig. 3 (Ch. 15, p. 196). M-FISH analysis of a metaphase obtained from the bone marrow of MDS patient.*

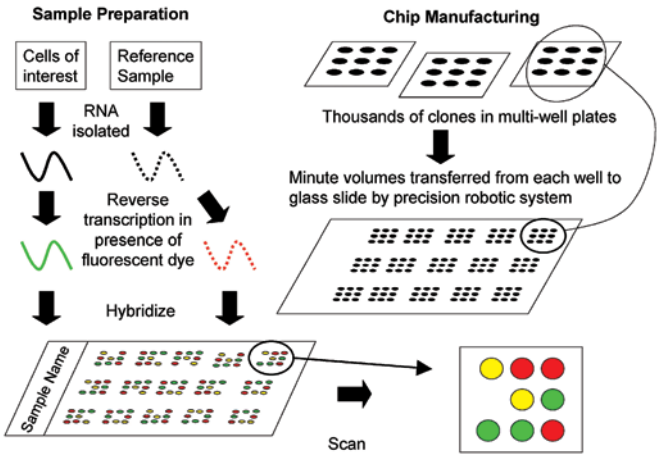
Color Plate 7: *Fig. 4 (Ch. 15, p. 197). Tyramide amplification of FISH signals.*

Color Plate 8: *Fig. 1 (Ch. 31, pp. 379–380). Examples of DNA mismatch repair immunohistochemistry.*

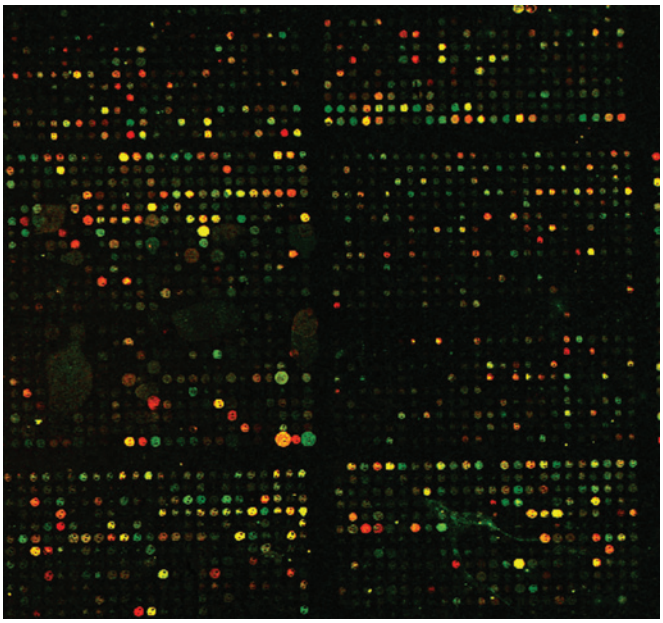
Color Plate 9: *Fig. 3 (Ch. 31, p. 383). Examples of microsatellite instability.*

Color Plate 10: *Fig. 4 (Ch. 31, p. 385). Comparative sequence analysis.*

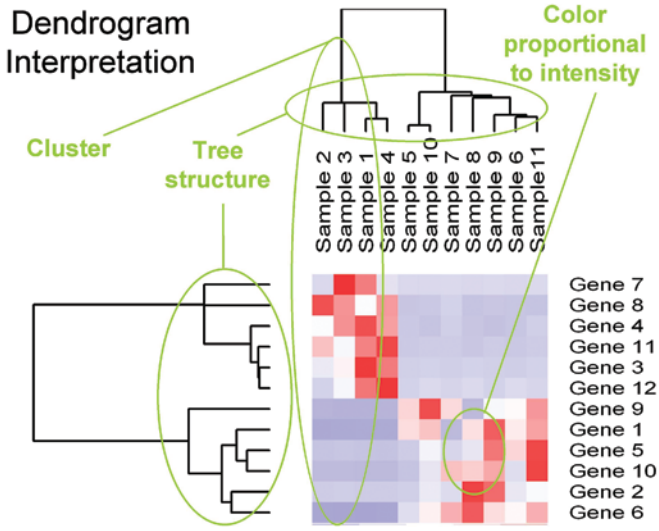
Color Plate 11: *Fig. 15 (Ch. 32, p. 404). FISH analysis utilizing a dual-color translocation probe system.*



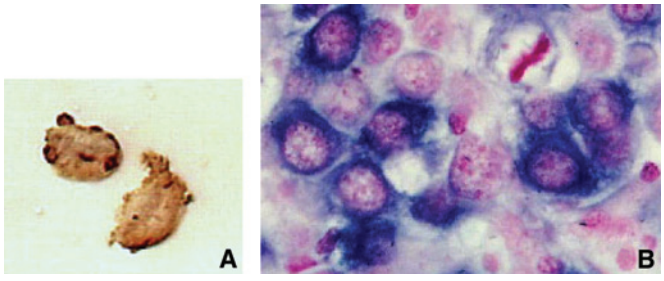
Color Plate 1, Fig. 1. (*See* full caption and discussion in Ch. 11, p. 124). cDNA microarray manufacturing process.



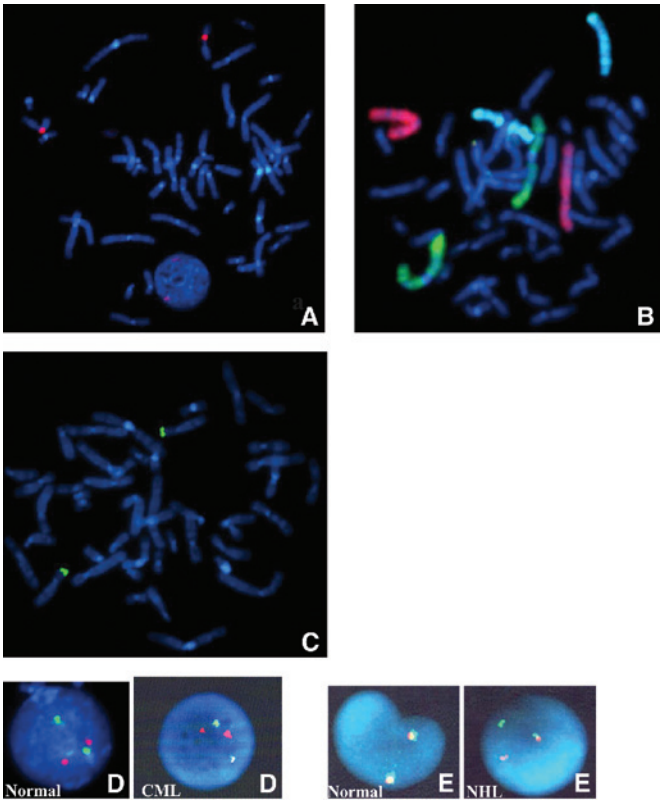
Color Plate 2, Fig. 2. (*See* full caption and discussion in Ch. 11, p. 125). Example of cDNA microarray image file. Magnified view showing approx 10% of the "spots" from a selected cDNA microarray.



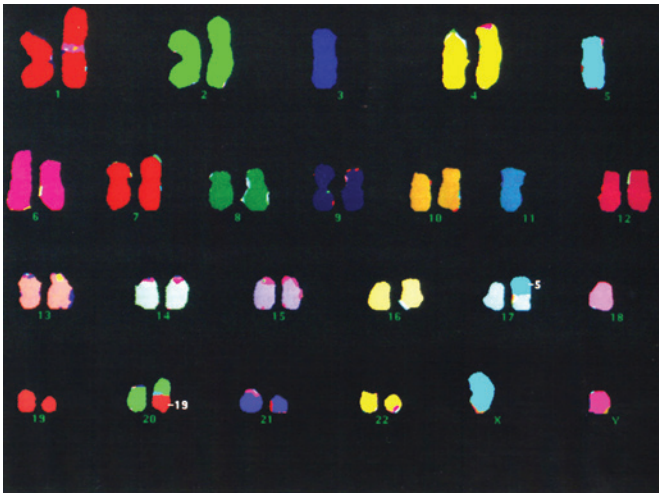
Color Plate 3, Fig. 11. (*See* full caption and discussion in Ch. 11, p. 138). Dendrogram.



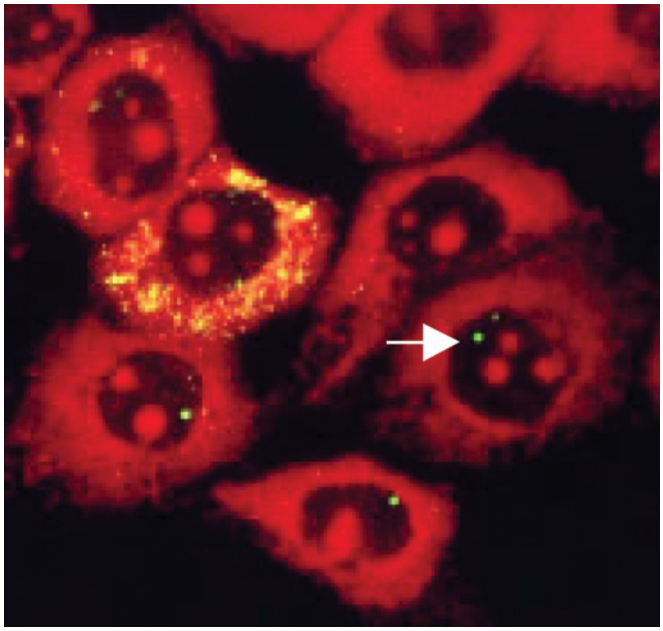
Color Plate 4, Fig. 1. (*See* full caption and discussion in Ch. 15, p. 191). *In situ* hybridization for detection of mRNA.



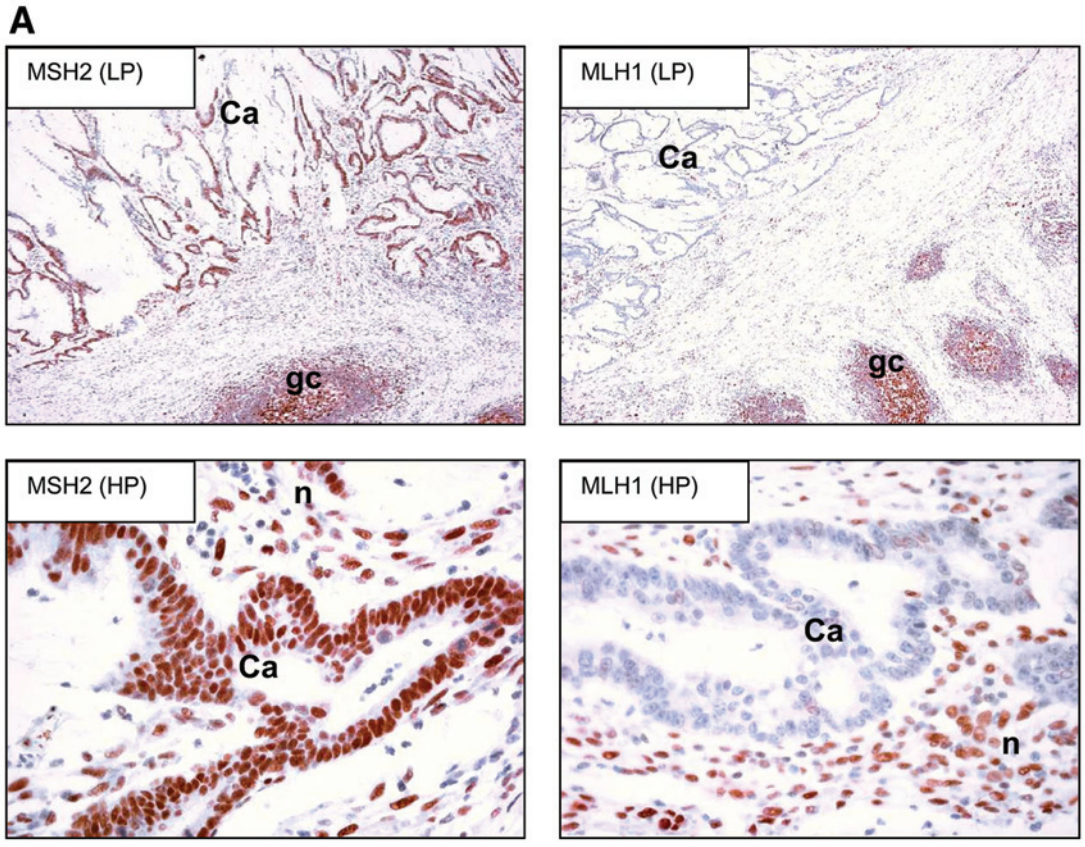
Color Plate 5, Fig. 2. (*See* full caption and discussion in Ch. 15, p. 194). Different types of DNA FISH probe.



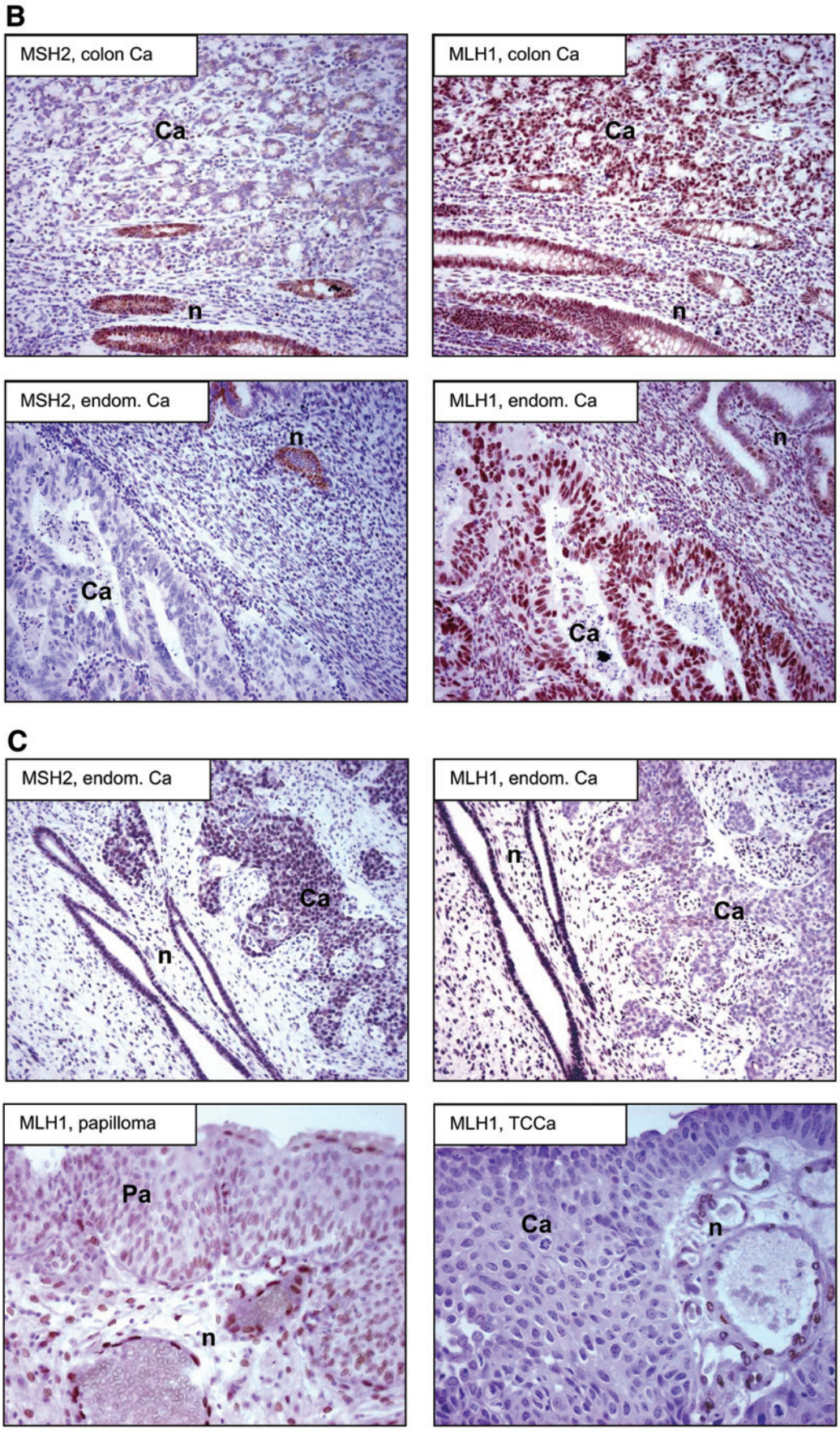
Color Plate 6, Fig. 3. (*See* full caption and discussion in Ch. 15, p. 196). M-FISH analysis of a metaphase obtained from a bone marrow of MDS patient.



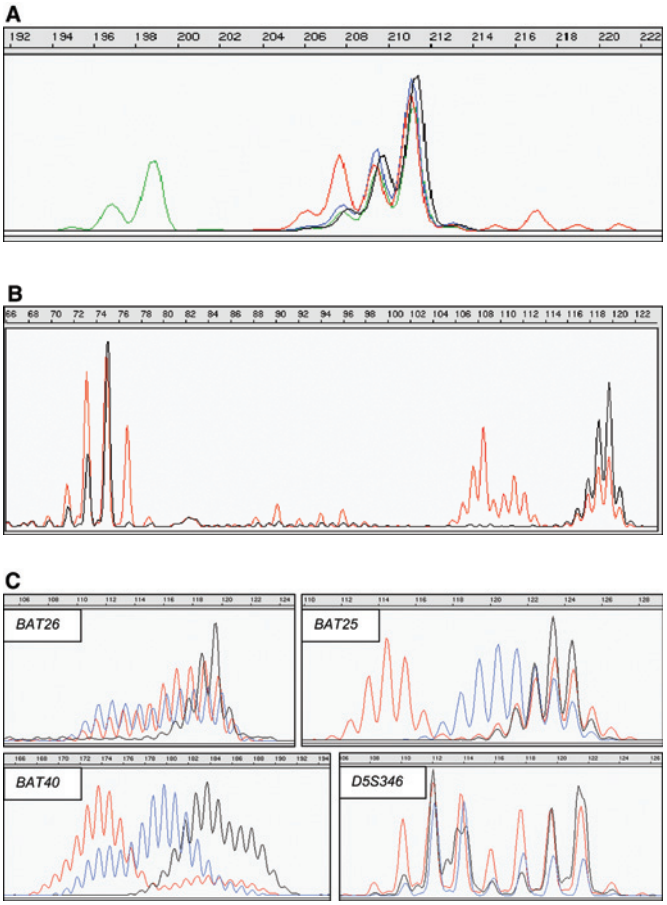
Color Plate 7, Fig. 4. (*See* full caption and discussion in Ch. 15, p. 197). Tyramide amplification of FISH signals.



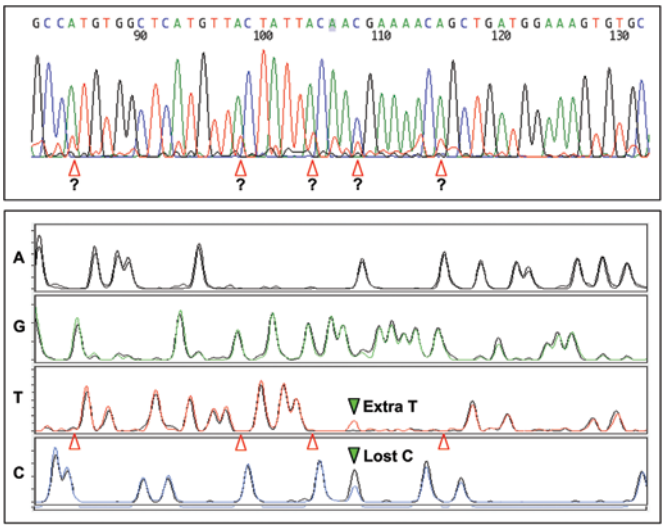
Color Plate 8, Fig. 1. (*See* full caption and discussion in Ch. 31, p. 379). Examples of DNA mismatch repair immunohistochemistry.



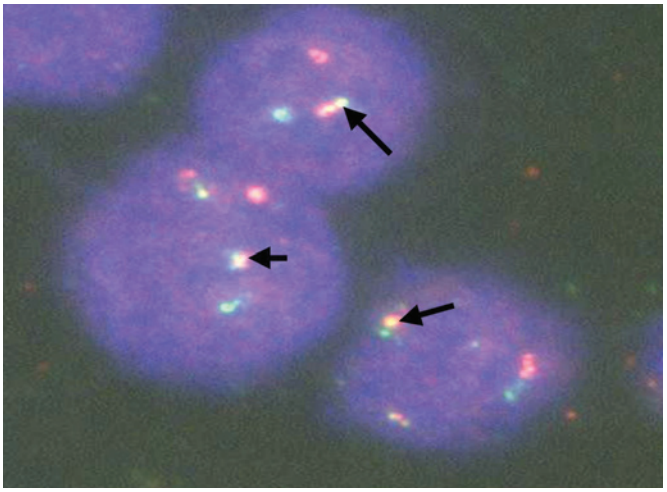
Color Plate 8, Fig. 1. *(Continued)*



Color Plate 9, Fig. 3. (*See* full caption and discussion in Ch. 31, p. 383). Examples of microsatellite instability.



Color Plate 10, Fig. 4. (*See* full caption and discussion in Ch. 31, p. 385). Comparative sequence analysis.



Color Plate 11, Fig. 15. (*See* full caption and discussion in Ch. 32, p. 404). FISH analysis utilizing a dual-color translocation probe system.

CONTRIBUTORS

TERRY AMISS, PhD • *BD Technologies/Biosensors, Research Triangle Park, NC*

JEAN A. AMOS, PhD • *Scientific Director, Molecular Genetics, Specialty Laboratorie, Inc., Santa Monica, CA*

MARK J. ARENDS, MA, MB ChB(Hons), BSc(Hons), PhD, FRCPath • *Department of Pathology, Addenbrooke's Hospital, University of Cambridge, Cambridge, UK*

C. ROBERT BAGNELL, JR., PhD • *Department of Pathology and Laboratory Medicine, University of North Carolina School of Medicine, Chapel Hill, NC*

MARIA BAIULESCU, MD • *Department of Pathology and Laboratory Medicine, Charlotte-Hungerford Hospital, Torrington, CT*

ENRIQUE BALLESTEROS, MD • *Department of Pathology and Laboratory Medicine, Hartford Hospital, Hartford, CT*

SELWYN J. BAPTIST, MD • *Department of Pathology, Saint Barnabas Medical Center, Livingston, NJ*

PAUL N. BOGNER, MD • *Department of Pathology, University of Michigan, Ann Arbor, MI*

JESSICA BOOKER, PhD, FACMG • *Department of Pathology and Laboratory Medicine, University of North Carolina School of Medicine, Chapel Hill, NC*

ULRICH BROECKEL, MD • *Department of Medicine and Human and Molecular Genetics Center, Medical College of Wisconsin, Milwaukee, WI*

RACHEL BUTLER, BSc(Hons), MRCPath • *Molecular Genetics Laboratory, Institute of Medical Genetics, Cardiff, UK*

ANGELA M. CALIENDO, MD, PhD • *Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, GA*

RICHARD W. CARTUN, PhD • *Department of Pathology and Laboratory Medicine, Hartford Hospital, Hartford, CT*

WILLIAM B. COLEMAN, PhD • *Department of Pathology and Laboratory Medicine, University of North Carolina School of Medicine, Chapel Hill, NC*

TRACEY DAWSON CRUZ, PhD • *Forensic Science Program & Department of Biology, Virginia Commonwealth University, Richmond, VA*

BRIAN R. CURTIS, MS, MT (ASCP)SBB • *Platelet and Neutrophil Laboratory, The Blood Center of Southeastern Wisconsin, Milwaukee, WI*

D. BRIAN DAWSON, PhD • *Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN*

LARISSA DAVYDOVA, MD • *Department of Pathology and Laboratory Medicine, Hartford Hospital, Hartford, CT*

JOSEPH A. DIGIUSEPPE, MD, PhD • *Director, Special Hematology Laboratory, Acting Director, Molecular Pathology Laboratory, Department of Pathology and Laboratory Medicine, Hartford Hospital, Hartford, CT*

ALEXANDER DOBROVIC, PhD • *Department of Pathology, Peter MacCallum Cancer Centre, Melbourne, Australia*

D. ROBERT DUFOUR, MD • *Pathology and Laboratory Medicine Service, Veterans Affairs Medical Center, Washington, DC*

ANDREA FERREIRA-GONZALEZ, PhD • *Department of Pathology, Virginia Commonwealth University, Richmond, VA*

IAN M. FRAYLING, MA, MB, BChir, PhD, MRCPath • *Clinical Genetics Laboratory, Institute of Medical Genetics, University Hospital of Wales, Cardiff, UK*

CARLETON T. GARRETT, MD, PhD • *Department of Pathology, Virginia Commonwealth University, Richmond, VA*

LISA HAPPERFIELD, FIBMS, MSc • *Department of Histopathology, Addenbrooke's Hospital, Cambridge, UK*

DAVID NEIL HAYES, MD, MPH • *Division of Hematology/Oncology, Department of Internal Medicine, Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC*

JENNIFER L. HERBST, JD, M.Bioethics • *Montgomery, McCracken, Walker & Rhoads, LLC, Philadelphia, PA*

MARTIN J. HESSNER, PhD • *Department of Pediatrics and Human and Molecular Genetics Center, Medical College of Wisconsin, Milwaukee, WI*

W. EDWARD HIGHSMITH, JR., PhD • *Division of Laboratory Genetics, Department of Laboratory Medicine Pathology, Mayo Clinic, Rochester, MN*

CHARLES E. HILL, MD, PhD • *Department of Pathology and Laboratory Medicine, Emory University Hospital, Atlanta, GA*

JOHN HUNT, MD • *Department of Pathology, Baystate Medical Center, Springfield, MA*

JOHN P. JAKUPCIAK, PhD • *DNA Technologies Group, Biotechnology Division, National Institute of Standards and Technology, Gaithersburg, MD*

MARTHA B. KEAGLE, MEd, CLSp(CG) • *Diagnostic Cytogenetic Sciences, School of Allied Health, University of Connecticut, Storrs, CT*

ANTHONY A. KILLEEN, MD, PhD • *Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN*

CHRISTOPHER J. KLEIN, MD • *Department of Neurology, Mayo Clinic, Rochester, MN*

ROBERT J. KOSKA, BSMT (ASCP), MBA • *Cepheid, Sunnyvale, CA*

EUGENE H. LEWIS III, DO • *Department of Pathology and Laboratory Medicine, Hartford Hospital, Hartford, CT*

MYRA J. LEWIS, DO • *Department of Pathology and Laboratory Medicine, Hartford Hospital, Hartford, CT*

PEI LIN, MD • *Department of Hematopathology, University of Texas M.D. Anderson Cancer Center, Houston, TX*

NELL S. LURAIN, PhD • *Department of Immunology/ Microbiology, Rush-Presbyterian-St. Luke's Medical Center, Chicago, IL*

CHRISTOPHER MATTOCKS, BSc(Hons) • *National Genetics Reference Laboratory, Salisbury District Hospital, Salisbury, UK*

KAREN MCCULLOUGH, PhD • *Vanda Pharmaceuticals, Rockville, MD*

L. JEFFREY MEDEIROS, MD • *Department of Hematopathology, University of Texas M.D. Anderson Cancer Center, Houston, TX*

JON F. MERZ • *Center for Bioethics, and Center for Clinical Epidemiology and Biostatistics, Department of Medical Ethics, University of Pennsylvania School of Medicine, Philadelphia PA.*

MATTHEW MEYERSON, MD, PhD • *Division of Oncology, Dana Farber Cancer Institute, Harvard Medical School, Boston, MA*

EMMA MONK, BSc • *East Anglian Regional Genetics Service, Molecular Genetics Laboratory, Addenbrooke's Hospital, Cambridge, UK*

ROBERT E. MOORE, PhD • *Department of Pathology and Laboratory Medicine, Hartford Hospital, Hartford, CT*

SUNEEL D. MUNDLE, PhD • *Rush Cancer Institute, Chicago, IL; Technical Services Manager, Vysis Inc., Downers Grove, IL*

NARASIMHAN NAGAN, PhD • *Laboratory Director, Athena Diagnostics, Worcester, MA*

KIM OAKHILL, BSc(Hons), DipRCP • *East Anglian Regional Genetics Service, Molecular Genetics Laboratory, Addenbrooke's Hospital, Cambridge, UK*

CATHERINE D. O'CONNELL, PhD • *Tetracore Inc., Gaithersburg, MD*

DEBORAH A. PAYNE, PhD, CLSp(MB), CLSDir(NCA) • *Department of Pathology and Otolaryngology, Division of Molecular Diagnostics, The University of Texas Medical Branch, Galveston, TX*

JOHN A. PHILLIPS III, MD • *Division of Medical Genetics, Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, TN*

AMY POTTER, MD • *Division of Endocrinology and Metabolism, Departments of Pediatrics and Medicine, Vanderbilt University School of Medicine, Nashville, TN*

SHARON COLLINS PRESNELL, PhD • *BD Technologies/ Biotherapy, Research Triangle Park, NC*

TERESITA CUYEGKENG REDONDO, MD • *Department of Pathology, Saint Barnabas Medical Center, Livingston, NJ*

WILLIAM N. REZUKE JR., MD • *Department of Pathology and Laboratory Medicine, Hartford Hospital, Hartford, CT*

ANDREW RICCI JR., MD • *Department of Pathology and Laboratory Medicine, Hartford Hospital, Hartford, CT*

SHARON L. RICKETTS • *Department of Pathology and Laboratory Medicine, University of North Carolina School of Medicine, Chapel Hill, NC*

MYRA I. ROCHE, MS, CGC • *Department of Pediatrics, University of North Carolina School of Medicine, Chapel Hill, NC*

MARLENE SABBATH-SOLITARE, PhD • *Department of Pathology, Saint Barnabas Medical Center, Livingston, NJ*

JOHN L. SCHMITZ, PhD • *Department of Pathology and Laboratory Medicine, University of North Carolina School of Medicine, Chapel Hill, NC*

BRENT L. SEATON, PhD • *Focus Technologies Inc., Cypress, CA*

LAWRENCE M. SILVERMAN, PhD • *Department of Pathology and Laboratory Medicine, University of Virginia, Charlottesville VA*

LAURIE E. SOWER, PhD • *Department of Microbiology and Immunology, The University of Texas Medical Branch, Galveston, TX*

STEPHEN N. THIBODEAU, PhD • *Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN*

GREGORY J. TSONGALIS, PhD • *Department of Pathology, Dartmouth Medical School, Dartmouth-Hitchcock Medical Center, Lebanon, NH*

ALEXANDRA VALSAMAKIS, MD, PhD • *Department of Pathology, Division of Microbiology, Johns Hopkins University School of Medicine, Baltimore, MD*

THERESA M. VOYTEK, MD • *Department of Pathology and Laboratory Medicine, Hartford Hospital, Hartford, CT*

ELAINE WEIDENHAMMER, PhD • *Senior Manager, Strategic Development, Nanogen, San Diego, CA*

MYRA J. WICK, PhD • *Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN*

ALAN H. B. WU, PhD • *Department of Laboratory Medicine, University of California, San Francisco, CA*

VALUE-ADDED EBOOK/PDA ON CD-ROM

This book is accompanied by a value-added CD-ROM that contains an eBook version of the volume you have just purchased. This eBook can be viewed on your computer, and you can synchronize it to your PDA for viewing on your handheld device. The eBook enables you to view this volume on only one computer and PDA. Once the eBook is installed on your computer, you cannot download, install, or e-mail it to another computer; it resides solely with the computer to which it is installed. The license provided is for only one computer. The eBook can only be read using Adobe® Reader® 6.0 software, which is available free from Adobe Systems Incorporated at www.Adobe.com. You may also view the eBook on your PDA using the Adobe® PDA Reader® software that is also available free from Adobe.com.

You must follow a simple procedure when you install the eBook/PDA that will require you to connect to the Humana Press website in order to receive your license. Please read and follow the instructions below:

1. Download and install Adobe® Reader® 6.0 software
   You can obtain a free copy of the Adobe® Reader® 6.0 software at www.adobe.com
   *Note: If you already have the Adobe® Reader® 6.0 software installed, you do not need to reinstall it.
2. Launch Adobe® Reader® 6.0 software
3. Install eBook: Insert your eBook CD into your CD-ROM drive
   PC: Click on the "Start" button, then click on "Run"
   At the prompt, type "d:\ebookinstall.pdf" and click "OK"
   *Note: If your CD-ROM drive letter is something other than d: change the above command accordingly.
   MAC: Double click on the "eBook CD" that you will see mounted on your desktop. Double click "ebookinstall.pdf"
4. Adobe® Reader® 6.0 software will open and you will receive the message "This document is protected by Adobe DRM" Click "OK"
   *Note: If you have not already activated the Adobe® Reader® 6.0 software, you will be prompted to do so. Simply follow the directions to activate and continue installation.
   Your web browser will open and you will be taken to the Humana Press eBook registration page. Follow the instructions on that page to complete installation. You will need the serial number located on the sticker sealing the envelope containing the CD-ROM.

If you require assistance during the installation, or you would like more information regarding your eBook and PDA installation, please refer to the eBookManual.pdf located on your cd. If you need further assistance, contact Humana Press eBook Support by e-mail at ebooksupport@humanapr.com or by phone at 973-256-1699.

*Adobe and Reader are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States and/or other countries.

INTRODUCTION 1

1 An Historical Perspective on the Clinical Diagnostic Laboratory

ROBERT E. MOORE

1. INTRODUCTION

As the clinical diagnostic laboratory has entered the 21st century, it is interesting to reflect on past scientific and social events that have influenced the status of the laboratory and to anticipate future problems and opportunities. It would be a mistake to suggest that all of the significant medical and social events that had an impact on the present laboratory function can be discussed or evaluated in a short introductory chapter. It is also possible to mistakenly attribute more influence to some events than they deserve simply because they are being interpreted with a 21st century bias. However, some threads of commonality have influenced the evolution of diagnostic laboratories from earliest times. It is the objective of this chapter to highlight events that, in my opinion, have shaped how and/or why clinical laboratories have arrived at their present position in the practice of medicine.

A history of diagnostic testing can be started by reviewing the evolution of diagnostic tests from isolated procedures to organized diagnostic laboratory testing. Originally, laboratory tests were performed at the side of the patient with small, simple equipment, rapid evaluation of the result, and a diagnostic opinion rendered. Test choice, performance, and interpretation were all left to the individual practitioner. There was no professional support staff to assist at any point in the process. Little exchange took place between practitioners because the individual's success was directly related to which procedures were done and the manner in which they were done. A premium was paid for showmanship as well as successful treatments.

The modern laboratory is a physical place, either standing alone or as a component of a health care institution, with numerous pieces of complicated capital equipment, where hundreds and sometimes thousands of specimens per day are processed for dozens of tests. Usually, the laboratory accepts a teaching and a research responsibility to accompany this patient service obligation. Today's laboratories are staffed by professionals trained in several subspecialties, who are available for consultation in all aspects of diagnostic testing. This structure underlines the current complexity and sophistication of the modern laboratory operation. There are a myriad of professional organizations, meetings, and publications that have developed to make information exchange convenient and efficient. This exchange is also necessary to ensure a single level of care. The system that supported showmanship and theatrics has been replaced by one that only tolerates scientifically based medical practices.

Today's laboratory is in a state of flux—faced with issues of near-patient testing, outreach programs, utilization, extensive regulation, and stringent fiscal controls. The first three of these issues are reminiscent of the early beginnings of diagnostic testing. The laboratory organization has replaced the individual practitioner, but testing is moving back toward the patient in the form of near-patient testing or outreach programs. In the past, a few individuals spoke about appropriate use of diagnostic testing, whereas, today, committees and organizations are dedicated to the control of laboratory utilization—an ontogeny recapitulating phylogeny phenomenon. A review of history gives some insight into relationships and similarities of past activity with current practice. What are perceived as new problems and opportunities can be traced to antiquity.

The development of the modern laboratory required several conditions to be met at appropriate times in history. Obviously, technology was and is the primary force behind advances in medicine. This has been both a blessing and a curse. It is a blessing because the understanding and treatment of disease requires sophisticated tools that only technological advances can produce. It can be a curse because there is a direct relationship between sophistication and health care cost. Other social issues dealing with availability and ethics assume a much greater importance than in the past. Angiography, computer-aided tomography (CAT) scans, organ transplantation, and DNA analysis are just a few of the expensive but valuable technologies available to the modern physician. Such questions as when are they used, to whom are they available, and what is done with the result are dilemmas of the 21st-century laboratory.

Another result of current technology is that it removes a large number of tests and procedures from the primary physician. The expertise required to perform these and other

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

procedures combined with the significant cost of the equipment precludes the primary physician from being the laboratorian.

A second condition that had to be met was logistics. Laboratory testing had to be convenient. All the concerns of collection, preservation, testing, and reporting had to be easy and fast for both patient and physician. A practical consequence of this was the proliferation of laboratories with increased resources to deliver the service. In the beginning, these resources were human and consisted of having specimens transported to the primary laboratory site. This later became multiple collection sites, courier services, mechanical processing, and electronic reporting.

These logistical issues led directly to a third condition for laboratory development: economics. There are two components to economic considerations. First, the laboratory service has to be an economically viable option for the ordering physician. The service must be delivered in a way so the physician does not experience any cost or significant loss of income from referring tests to a diagnostic laboratory. Second, there must be a mechanism to support the cost of laboratory testing. The latter problem was solved, for a short time, by the third-party reimbursement system. As one of the social reform programs of the 20th century, the widespread availability of insurance made the cost of health services invisible to the patient. Without proper controls, the effect of this reimbursement system was to encourage the proliferation of technology and make services accessible to large segments of the population. The incentive was for every testing center to have all the best technology and make it available to everyone.

This chapter will highlight some historical events and practices that demonstrate how this evolution took place. The events outlined here are not absolute in defining the practice of laboratory medicine, but they suggest how concepts could evolve and develop into the practice of diagnostic laboratory testing as it is today, including some events that could be precursors of molecular diagnostics.

2. EARLIEST MEDICINE

Long before there were laboratories, there were accepted practices for patient evaluation. Early health care providers (not all were physicians) attempted to determine the health status of the person under evaluation by any means possible. The diagnosis and the prescribed treatment were not always an accurate or scientifically based pronouncement. The process was motivated by a combination of altruism, vanity, greed, scientific thought, and philosophical and religious edict. This is not meant to imply that all was quackery and incantation. The procedures that had medical value are the ones that laid the foundation for legitimizing diagnostic medicine along with its subsequent support functions, one of which is the clinical laboratory.

One major obstacle that physicians faced in ancient times was that it was illegal to practice invasive procedures. The patient could be observed and touched, but the only specimens that could be taken were those that naturally passed from the body. Because of these limitations, urine has been the sample with the longest history of evaluation. There is some evidence that the Sumerians and Babylonians used urine for diagnosis as early as 4000 BCE *(1)*. The diagnosis of pregnancy was probably made by ancient Egyptians using the urine of the woman to germinate seeds *(1)*. Hindu medicine describes the sweet taste of urine and that black ants are attracted to this urine if it is poured on the ground *(2)*. Hippocrates (460–355 BCE) described the characteristics and colors of urine from his patients and mentioned bubbles being present in urine from patients with long-standing kidney disease *(2,3)*. Over the next 600 years, the study of urine was advanced very little. Galen (129–200 AD) wrote and taught that urine was a filtrate of blood and, as such, could indicate the type and location of illnesses *(2)*. The teachings of these two men were the information base for urinalysis, or uroscopy as it was called, for the next 9–10 centuries.

During this time, approximately 800 AD, the first treatise on urine was written by Theophilus Protospatharius; in it is mentioned the first chemical test done on urine. Urine collected from patients with kidney disease was heated over a candle flame and became cloudy *(2)*. Other physicians repeated the process, some substituting acid for heat, and although it took centuries before the precipitating substance was identified as protein, the association with disease had been recognized *(3)*.

Other observations were made concerning the quality of the urine sample. The differences between morning and afternoon urine samples and factors like age, food, and medicaments exerted effects on the composition of urine. These were noted as early as the 10th century by Avacinna *(2)*. It was realized as early as the 11th century that the first voided specimen in the morning was the best urine specimen for analysis and that when 24-h urine collections were required, they should be protected from light and heat *(2)*.

Urinalysis continued to be such a focus of study that Gilles de Corbeil developed a glass vessel (called a matula), shaped like a urinary bladder, specifically for the examination of urine. The concept was that the urine sediments and discoloration would occur in the vessel at a place that corresponded to the site of pathology in the body. These early urine vessels were among the first pieces of laboratory equipment and were so common and identifiable that they were one of the predecessors to the caduceus as the symbol of medicine.

Other tests and procedures were added to urinalysis and various aids were developed to make the process easier. One device of this type was the urine wheel. This wheel was the original color chart similar in purpose to those that accompany most modern dipstick packages. Along with matching the urine color on the wheel, there was an interpretive text included to assist in making a diagnosis *(3)*.

Uroscopy soon took on a life of its own. Samples were sent to physicians without any explanation of the patient's complaint and it was expected that the analyzing practitioner would return a diagnosis and therapy *(3)*. Expectedly, uroscopy was ripe for abuse and, assuredly, this happened. The prominence of the practitioner was enhanced if he was perceived as being able to do more than other uroscopists. Consequently, claims were made concerning the interpretive powers of the analyst that far exceeded the limits of observation. It was during this period, when there was an opportunity to make a handsome income from urine analysis, that Joannes Actuarius began to write about the limitations of urine examination. He was one of the first to caution that urine examination, independent of how well

it was done, could not be used to the exclusion of all other clinical findings *(3)*.

It should not be inferred that uroscopy was always a questionable effort. Proteinuria, although protein as such had not been identified, nephritis, type I diabetes, hematuria, infection, concentration, and limited assessment of liver disease were all recognized through urinalysis. In most cases, it took centuries to identify the specific component in urine and its association with a disease; nevertheless, urine testing was a valuable diagnostic tool *(3)*.

These early attempts to diagnose disease, through the study of the only practical body fluid available, gives some insight into the development of laboratory medicine. First, the progression from pure observation to the use of some elementary aids and procedures suggests that the seed of intellectual curiosity was germinating. If simple observation was useful, then employing procedures to define components was better. Second, equipment was being generated for a specific medical application. The equipment was primarily small, and portable but with such procedures as distillation, precipitation and evaporation, the need for a place to do this work was becoming an issue. A permanent site or address for the practitioner to do his laboratory work enabled patients to "send" their urine for analysis (an early precursor to the outpatient laboratory). Third, because uroscopy was becoming a routine practice, some members of the legitimate medical community were discussing the appropriate use of these procedures. This might be the first suggestion of test utilization in history.

3. THE TRANSITION

Progress requires an advance in technology or the appearance of a gifted individual. Science and technology had advanced to a point where a breakthrough on some other front was required. Two historic events that meet this requirement are the invention of movable type by Johann Gutenberg in the 15th century and the Reformation. Movable type and the printing press had the potential to make information available to large numbers of scholars. Experiments and data could be corroborated by other professionals, and groundless claims could become less common. Education, in all its forms, would be a benefactor. Education needs an appropriate environment. The Reformation provided that environment. The individual was acquiring more freedom in life, and the old way or restricted and regimented thought was passing. In this time of change, there were individuals of the proper education and personality to take advantage of the situation.

Andreas Vesalius, who is credited as the founder of modern anatomy, was such an individual. He was the first to do extensive dissections and accurately document the anatomical information in text and illustration *(4)*. The importance of this lies in the transition from animal to human dissections. Accurate anatomical data would become available and true progress in the other anatomical sciences, histology and zoology, could take place *(4)*. Another individual of this type was Giovanni Battista Morgagni, a successor to Vesalius, who combined anatomic dissection with clinical history to describe pathologic anatomy *(4)*. This was arguably the beginning of modern pathology.

It was during the 17th century that the science of laboratory medicine began to acquire the characteristics that are easily recognized today. Laboratories were being established in homes in which research and development was being conducted *(4)*. The results of this cottage industry were the development of equipment for laboratory and diagnostic use. Thermometers and hydrometers were two such pieces of equipment *(4)*. Although the thermometer was described by Galileo in the late 1500s, it was not used for patient evaluation until the 17th century. The arguments waged as to which fluid was the best indicator of change and the best way to calibrate the instrument *(4)*. Given these and other practical problems, some physicians were quick to realize that a quantitative measure of body temperature was an improvement over using one's hand. For other physicians, the problems with obtaining and using the thermometer were too great. After this brief flurry of activity, the thermometer returned to a quiescent state until the 1800s. By the 1800s, many of the original problems had been either resolved or minimized, so it became a more practical matter to measure temperature. Articles appeared regarding the value of temperature in diagnosis and prognosis in major medical journals *(5)*.

The hydrometer had relatively little difficulty in its acceptance. The instruments were made small and reliable, so they could be taken on home visits to the sick. Specific gravity was one of the well-entrenched measurements of uroscopy. A reliable and simple instrument catered to the needs of the common practitioner. Urinalysis was still king.

The home lab contributed to the refinement of equipment and methods. It should be remembered that alchemy was slowly giving way to organic chemistry, which was defined as the chemistry of living organisms *(6)*. It was not until the synthesis of urea in 1828 that organic chemistry became the chemistry of carbon compounds. Consequently, the efforts of the more inquisitive chemist-physicians turned to chemical procedures on body fluids *(6)*. Paracelsus and, later, Willis suggested chemical analysis of urine *(6)*. To separate the fluid into its component parts was thought to yield better diagnostic or prognostic information *(6)*. True to history, a new hypothesis always has its detractors. Perseverance by the investigators and positive, predictable results from the applications generated broader acceptance.

It was an extension of this philosophy of active investigation rather than passive observation that led Robert Boyle to analyze blood. He had done chemical analysis on urine and continued this approach on human blood *(5,6)*. Boyle had the disadvantage of not being a physician and so did not have the opportunity to analyze the blood of sick patients *(5)*. He did have the insight to propose that knowing the results of his chemical analysis on healthy people, the evaluation of the sick would be made easier *(5)*. Boyle was obviously an early proponent of normal reference intervals. This concept was carried further by the French physician Raymond Vieussens, who chemically analyzed the blood of large numbers of people of all descriptions *(5)*. Normals, abnormals, men, women, and those of *different temperaments* were included in his study. These studies of defined populations were exceptional in their time.

Diagnostic testing had achieved a credibility that made it acceptable to perform chemical analysis on blood and urine and

use the information in a diagnostic workup. Laboratories were beginning to grow in number and they were developing techniques that could be applied to medicine. Laboratory skills were gaining respectability along with laboratory testing and laboratory data.

4. EARLY MODERN MEDICINE

Chemistry as a science was developing rapidly in the 18th and early 19th centuries. Diabetes was the subject of intensive investigation by the medical chemists. Physicians like John Rollo and William Cruikshank demonstrated the presence or absence of sugar in the urine of diabetics, depending on the state of the patient's disease *(5)*. They were proponents of being able to monitor a patient's status by knowing if there was sugar in the patient's urine *(5)*. The conventional wisdom of the time was that most of the practicing physicians were not sophisticated enough to do this type of analysis routinely *(5)*. However, the concept of monitoring through repeated testing was being considered.

In 1827, Richard Bright published his *Reports of Medical Cases,* in which he described albumin in the urine of patients with dropsy, which is the collection of serous fluid in body cavities or cellular tissue. Bright was aware of the anatomical lesions that were associated with this condition and he made the link between chemical findings and anatomic findings *(5,7)*. This condition was later referred to as Bright's disease in honor of the association made by Bright *(7)*. This undoubtedly gave considerable credence to the analysis of specimens to diagnose disease. The test was simple: Heat the urine and observe the white precipitate that formed. The presence of kidney disease could be predicted by the outcome of the test. The test was not perfect, and as more people performed the test, conflicting results were obtained. Some "albuminous" material did not precipitate unless chemicals were added, and some normal people had precipitates in their urine *(5)*.

These and other arguments were responsible for the slow acceptance of Bright's albumin procedure and its use in diagnostic medicine. However, the principle of laboratory testing and the additional insight that it gave the physician was developing slowly. Progress was on the front side of a geometric curve, and by the end of the first half of the 19th century, greater utilization and faster development of laboratory diagnostic medicine would be commonplace.

A major driving force for this increased activity was the work of Gabrial Andral. Recognizing the value of chemical analysis of body fluids, he argued effectively for increased research by the medical community *(5,6)*. He was a believer in chemical and microscopic examination of the liquid as well as the solid components of the body *(5,6)*. His work, *Pathological Hematology,* examined blood chemically, microscopically, and visually in both the healthy and diseased populations *(6)*. He measured or calculated the major components of blood and was able to demonstrate a decreased red cell mass in anemia and a decreased blood albumin in albuminuria *(6)*. The success of Andral's work coupled with his enthusiasm led other workers to pursue similar investigations. The studies of sugar in the blood of diabetics and uric acid in the blood of those who suffered from gout are prime examples of this effort *(5)*.

The hematology specialties were benefiting from Andral's work. The invention of counting chambers for red cell quantitation and their use in diagnosing anemia was discussed in medical publications *(8)*. All the work in chemical analysis of fluids and the microscopic analysis of blood was facilitated because bloodletting was still an acceptable form of therapy. This was a Catch 22 situation. As more information was gained about diseases like anemia through these new methodologies, it became apparent that bloodletting was not appropriate therapy and this inexhaustible source of study material would become less available.

The last half of the 19th century was a prolific time for the development of laboratory methods. So many methods began to appear that the practicing physician had difficulty in choosing which assays were reliable and which were not *(8)*. Time and equipment were also becoming impediments to the routine use of these new tests. Physicians complained that the time required to conduct these analyses and the expertise involved in their performance exceeded their capabilities *(8)*. The problem would get worse before it ever got better. Chemistry was becoming a tool to be used by all the medical disciplines. Dr. Paul Ehrlich used chemical dyes to test urine in the diagnosis of typhoid fever and aniline dyes to distinguish different types of leukocytes *(7,8)*. The study of gastric disturbances was accomplished by chemical analysis of stomach contents removed by gastric tube *(8)*.

The medical literature was replete with discussions of new methods and of the failures and shortcomings of older ones. If technological advances were the only requirement for legitimacy, then laboratory medicine was an established discipline. However, true acceptance can only be established through academic credentials. Throughout the 19th century, there was a parallel development of the academic and political status of diagnostic laboratories. Initially, medical schools did not have the status of universities and it was well into the 19th century before they became equivalent to colleges *(9)*. Around the same time, medical chemistry was being split from chemistry, which was either general or organic chemistry *(9)*. The latter subjects were considered to be more pure science and, therefore, more prestigious. The result was that physiological chemistry was the domain of physiology with no academic standing of its own *(9)*. There were notable exceptions, one being the University of Tübingen, at which the first chair of physiological chemistry was established in 1845 *(9)*. The first chairman was Julius Schlossberger, who was responsible for all the chemistry teaching in the medical facility *(9)*. He was followed by Felix Hoppe-Seyler under whom the chair was transferred to the philosophical faculty *(9)*. The chair survived until the 20th century, primarily because of the stature of the chairmen, all of whom were both organic and physiologic chemists *(9)*. Similar academic structures were not as successful at other universities. It was not until the position was established in the medical schools of the United States that physiological chemistry became secure.

The first laboratory of physiological chemistry in the United States was established in 1874 at the Sheffield Scientific School of Yale University under the direction of Russell H. Chittenden *(10)*. This was rapidly followed by similar laboratories being

established at other major universities. Faculty members were expected to teach and conduct research, and professional positions were awarded on the academic credentials of the applicant *(10)*. These facilities were a permanent part of universities and, later, medical schools. The premedical and medical training of physicians included laboratory training in the biological and chemical sciences.

Another development in the late 1800s was the appearance of hospitals in the United States. By midcentury, these hospitals were designing laboratory space for purposes of urinalysis *(7)*. Resources were being made available for laboratory work, of which the overwhelming volume was urinalysis *(7)*. There is some indication that urinalysis was routine in at least one prestigious Eastern hospital *(7)*. At the end of the century, the hospital laboratory had been joined by the ward laboratory *(8)*—a smaller version of the main laboratory. The rationale was that a small lab space near patients would reduce the length of time required to get results and could be staffed by physicians and house staff. In practice, these two types of laboratory grew, each requiring more resources to make them operational.

Two physicians added considerable credence to the concept of a professionally staffed, hospital-based laboratory. Otto Folin, at a lecture in 1908, proposed that laboratories should be hospital based and staffed by professional physiological chemists *(11)*. William Osler judged the value of the laboratory to be indispensable to the clinical physician. When physicians as influential and respected as these were become proponents of laboratory testing, the position of the laboratory was permanently secured *(8)*.

With the continued acceptance of the laboratory and hospitals considering them to be integral to their service, a subtle change began to take place. Once the lab became established, work was generated. A review of several hospitals' records at the beginning of the 20th century indicates that urine testing was being done on most patients even if there were no indications for such a test *(7)*. These urine tests were being refined by investigators who now had positions in the laboratory. New tests were being added for blood and other body fluids. Many of the famous names associated with clinical laboratory medicine such as Folin, Benedict, Garrod, Koch, Van Slyke, and Ehrlich, among others, came from this era. This was a very productive time for research and development; many of these findings were being transferred to the diagnostic service laboratory.

Urine and diabetes have been studied for centuries in the hope of providing better care. There are records at the Pennsylvania Hospital that urine sugar measurements were used to monitor the therapy for a diabetic woman *(7)*. The end point was a negative urine sugar finding and so urine was analyzed every day. This was one of the first records of using lab tests to monitor treatment. After insulin was discovered, it became more important to monitor sugar because of the difficulty in controlling insulin therapy. Insulin preparations were of different purities, and external factors such as exercise and diet made insulin dosing very difficult. Blood sugar analysis was possible but difficult and so it became routine to do regular urine sugar measurements *(7)*.

Methods continued to be developed and clinical applications were tested. Each test or procedure found its way into the service offerings of the clinical lab. It was apparent that at least two problems were continuing to plague the lab. First, the volume of work was increasing and projections indicated that the trend would continue. Second, laboratory tests were difficult and tedious to do and, as a result, showed significant variability from imprecision. A partial solution to these issues came from Dr. Leonard Skeggs. His design of a continuous flow analyzer was the first practical unit for the laboratory *(12)*. These first designs were essentially mechanical duplicators of hand procedures but enabled the lab to increase throughput and improve the precision of the analysis. Automation for laboratory testing has undergone several generations of change. These instruments are now found in all sections of the laboratory and encompass a wide spectrum of methodologies.

5. THE MOLECULAR LINE

As the 21st century begins, the clinical diagnostic laboratory is entering into a new phase. Biotechnology, in all its forms, is the fastest growing discipline in the modern clinical laboratory. From the original experiments of Gregor Mendel in 1865 describing hybridization of plants, to the Human Genome Project, molecular biology is presenting the laboratory professional with new challenges. These challenges are not limited to the laboratory professional and questions of science. Ethical, legal, and commercial questions are every bit as daunting as the scientific issues.

Molecular biology appears to be a relatively recent discipline, but it is possible to point to events in the distant past that can be considered precursors of the current science. The Assyrians and Babylonians, between 5000 and 2000 BCE, recognized the existence of gender in the date palm tree and undertook artificial pollination *(13)*. It can be surmised that the purpose was to enhance those traits that were considered desirable. Between the 6th and 4th centuries BCE, the Greeks discussed inherited traits and the relative influence of inheritance versus the environment with respect to birth defects *(13)*. Arguments were put forth as to how humans developed and what was the source of the traits that were easily recognized between parent and offspring.

Progress was slow until Maupertius described the first inherited genetic disorder through four generations of a family in 1752 *(13)*. Joseph Adams published on the hereditary properties of disease, delayed expression of hereditary diseases, and environmental exposure as a trigger for disease *(13)*. This was quickly followed by Schleiden and Schwann proposing that nucleated cells were the fundamental units of life and Virchow's theory that cells could only come from the division of existing cells *(13)*. In 1859, Darwin published "On the Origin of the Species," a significant work that was lacking in genetic theory *(13)*.

The point in history where most people agree that molecular genetics began is 1865 when Gregor Mendel presented his work with peas to explain hereditary traits *(14)*. As important as this work is now viewed, it must be remembered that it went essentially unrecognized for 35 years. For the remainder of the 1800s, there were many discoveries and theories of cell physiology and structure. Some of the current terms in molecular biology and genetics such as nucleic acid, chromosomes, and mitosis come from this era.

The first half of the 20th century was fueled by Mendel's work. Traits in plants and animals were re-evaluated in light of the Mendalian ratio. Remarkable studies by remarkable scientists contributed a large volume of basic science to this burgeoning discipline. Such contributions as chromosomes are paired (1902), X and Y chromosomes determine gender (1905), genes are physically on chromosomes (1910), the first gene map (1913), and DNA isolated and purified (1935) were obligatory precursors to the Human Genome Project *(13,14)*.

There were other events during this period that have an interesting relationship with modern molecular biology. First, the stress of every first-year genetics student can be traced to 1910, at Columbia University, when T. H. Morgan determined the sex-linked character of some traits in *Drosophila melanogaster*. A choice of model that allowed the study of many generations in a short time span. This became important because Hermann Muller used the fruit fly model to demonstrate how ionizing radiation accelerated the formation of genetic mutations. This was important several years later to the Department of Energy. Second, X-ray crystallography studies of DNA and proteins led to the origination of the term *molecular biology*, the implication being that there had been a movement from the relatively gross studies of biology to a more refined study on a molecular level. Third, the concept of one gene, one enzyme was proposed by George Beadle and Edward Tatum *(15)*. This theory unified several of the discrete pieces of information that had preceded it and also unified genetics and biochemistry *(15)*. This is the first major discovery of molecular biology *(15)*.

Up until 1950, information was being gathered, unified, and converging to a point that is now called molecular biology. However, as the computer scientists say, there is another thread, one that is more closely related to the practice of current molecular biology. In the early part of the 20th century, Archibald Garrod published his findings about the hereditary nature of alkaptonuria *(15)*. The significant issue here was that he was able to show this to be a metabolic or chemical disorder *(15)*. Of the cases he studied, all were the product of a union between first-degree relatives, and Garrod was able to reason that this particular mating practice allowed a recessive character to surface, as predicted by a Mendelian distribution *(15)*. This finding was followed a few years later by his book *The Inborn Errors of Metabolism*. The exact chemistry of the involved metabolic pathways would not be known for several years, but this is the first instance of laboratory work yielding a diagnosis, similar to today's practice.

A similar discovery took place in 1934. Asbjorn Folling was able to relate mental retardation to the metabolic disorder phenylketonuria (PKU), an inborn error in the metabolism of phenylalanine. The error arises because of a mutant gene in the synthesis of the enzyme phenylalanine hydroxylase *(13)*. This represents another application of molecular biology; a genetic disease is detected very early in life and a treatment regimen can be started to prevent the disastrous sequelae.

In 1949, Linus Pauling and his research group published an article in the journal *Science* that described sickle cell anemia as a molecular disease *(15)*. This is the first description of a medical disease on a molecular basis and attributed to a mutant gene. Pauling followed this with a description of the α-helix structure of proteins *(14,15)*. This elucidation of the three-dimentional structures of proteins was a remarkable feat; yet Pauling did not discover the structure of DNA.

These efforts continued in the 1950s with the establishment of the number of human chromosomes at 46 and the discovery of the chromosome abnormalities Down syndrome (trisomy 21), Turner syndrome (45,X), and Klienfelter syndrome (47,XXY) *(13)*. This was an impressive start for the discipline of cytogenetics. Although these advances are some of the earliest representations of how medical molecular biology is practiced, the fundamental science was advancing at a remarkable rate. Chargaff determined that the ratio of the nucleic acid bases was always 1 : 1, an important breakthrough for determining the structure of DNA *(13,14)*. Combined with the crystallography work of Rosalind Franklin and Maurice Wilkins that showed an orderly, multiple polynucleotide chain, helix structure, James Watson and Francis Crick were able to propose a structure for DNA *(16)*.

Many pieces of the puzzle were being put in place by the extraordinary work of many groups. The gross structure and many technical details of the components of DNA were becoming available to the research world. One very important question about how this package of information was translated to actual proteins remained. Marshall Nirenberg and Heinrich Matthaei conducted a series of experiments with RNAs to see if they would synthesize proteins. From this series came the "poly U" discovery that UUU was the code for phenylalanine *(16)*. Along with significant contributions from Gobind Khorana (polynucleotide synthesis), Philip Leder (tRNA binding to ribosomes to determine code), and many others, this opened the way for the remainder of the code to be broken.

As important as the fundamental discoveries were, equally important developments were made in methods and technologies. This period saw the isolation and characterization of reverse transcriptase, DNA ligase, and restriction enzymes, methods for staining chromosomes, the Southern blot assay for DNA fragments, an approach to determining the nucleic acids in DNA, phage and plasmid development, and the polymerase chain reaction. The polymerase chain reaction was a sort of integration of all that had gone before. This relatively simple procedure made it possible to characterize DNA, even though the source material was of limited quantity. It was an active time that was making all of the tools available for cloning.

Cloning, as a process, was accomplished before the word was coined. In 1952, Robert Briggs and Thomas King used the technique to clone frogs *(17)*. The method was a nuclear transfer procedure that was later improved and replicated by John Gurdon at Cambridge *(17)*. There were other experiments with the fruit fly and bacteria, but in the early 1970s, Paul Berg created the first recombinant DNA and Stan Cohen, Herb Boyer, and colleagues created the first recombinant DNA organisms *(16)*. They had successfully amplified toad DNA in *Escherichia coli.* This was the beginning of genetic engineering and began the discussion of the social impact of this new science.

The ability to alter genes and to amplify them in another species was a source of concern to say the least. Could these new "agents" cause deadly diseases that would ravage mankind?

This was truly uncharted territory, and for the next few years, the scientific community discussed what, if anything, should be done to control the science. An international conference in February 1975, held in Pacific Grove, CA (Asilomar Conference), generated a set of provisional recommendations that were later used by the National Institutes of Health (NIH) to formulate a set of mandatory guidelines. All NIH-funded programs were obligated to follow these guidelines *(16)*. Other agencies around the world soon adopted similar restraints on recombinant DNA research.

An unusual situation was now developing. The potential of cloning was not lost on anyone and each experiment received wide exposure in both the popular press and scientific journals. Therefore, enterprising individuals saw the commercial possibilities and the early biotech companies were born. Genentech, Cetus, Genex, Biogen, and Amgen were some of the early entries into genetic engineering *(16)*. In 1980, it became legal to patent genetically engineered organisms. This encouraged the pharmaceutical and research companies to pursue protein hormones, drugs, and specific links between genetic abnormalities and diseases. Somatostatin was the first genetically engineered hormone and was followed in the next year by insulin and shortly thereafter by erythropoietin. It was clear that this was the new approach to the production of pharmaceuticals.

The cloning experiments that were conjecture in the 1950s and 1960s became reality in the 1980s. Nuclear transplant cloning of mammals was accomplished on mice, sheep, bovines, pigs, goats, and rabbits *(17)*. These experiments and others paved the way for the cloning of Dolly, the first animal cloned from adult cells. Dolly was born in July of 1996 at the Roslin Institute in Scotland *(17)*. The theories had been turned into practice and the pharmaceutical industry as well as all of biological science would embark on an exciting and daunting future.

One of the greatest success stories in molecular biology has to be the Human Genome Project. Some time after WWII, the Atomic Energy Commission (AEC), later known as the Department of Energy (DOE), was intensely interested in studying the health effects of ionizing radiation *(18)*. The AEC had been consumed with the Manhattan Project and the creation of the atomic bomb and one can reasonably assume that the interest in ionizing radiation was piqued by those events. The AEC was the largest funder of genetic research in the United States. By the 1980s, the DOE was supporting research on the health impact of non-nuclear sources of energy *(18)*. In this environment, Charles DeLisi began to muse about mapping the human genome. The obstacles were tremendous. Initially, it was thought that the process that would be required to accomplish such a feat was impossible. Techniques were not available yet to make this practical. Major agencies like the NIH were not interested and neither were many of the ranking scientists throughout the country. The idea persisted, and as technologies were developed and computer power was improved, there began to be more interest. The DOE announced in 1987 the formation of the Human Genome Project Initative that would order and sequence the human genome *(18)*. In science as in all of life, nothing becomes more interesting than when someone else has an interest. This announcement by the DOE stirred the interest of the NIH or, more specifically, the director, James Wingaarden *(18)*. The scientific community was still divided because the perception was that the money used for this project would jepordize all the other funded research. The cost of this project was estimated in the billions of dollars. The National Academy of Sciences wrote a report for the NIH that supported the genome project. The funding war was on. DOE and NIH were cosponsors of the project in the early days, but funding gradually shifted to the NIH *(18)*.

The official announcement of the Human Genome Project was made in 1990 and it was expected to take 15 years and 3 billion dollars. The original goals were to generate a high resolution gene and physical map of the human genome, to determine the complete DNA sequence in humans and other organisms, to develop the technology to store, analyze, and interpret the data, and to assess the ethical, legal, and social implications of genomics. It is interesting to note that the ethical, social, and legal issues were funded from the same source as the scientific project. Remarkable progress and cooperation, as evidenced by the sharing of sequence data within 24 h, was a hallmark of the project. That is not to say there was not some competition from the commercial side. Celera Corporation was created to compete with the NIH project and had set as its goal a 3-year timetable and a significantly reduced cost.

President Clinton announced in June 2000 the completion of the first draft of the Human Genome Project. It was a joint presentation, with both the commercial company Celera and the International Human Genome Consortium represented. At the time of the announcement, the project had cost about 300 million dollars and was several years early *(13)*. It was truly a first draft in that the entire genome had not been mapped, but by 2001, about half was finished and available in the public database *(13)*. Fifty years after the announcement of the structure of DNA, the Human Genome Project has accomplished virtually all of its goals.

From this brief overview, it can be said that molecular biology has developed, more than any other science, by the cooperative effort of many diverse disciplines. There was the mathematical approach of Mendel employing algebraic logic *(19)*. The isolation of protein-free nucleic acid by Richard Altman and the experiments to convert nonvirulent to virulent bacteria by Fred Griffith and Oswald Avery are significant developmental landmarks *(20)*. The application of the principles of theoretical chemistry by Max Delbrück and Erwin Schrödinger allowed others to proceed using the principles of physical science *(19)*. James Watson and Francis Crick elucidated the mathematically satisfying structure of DNA *(21)*. There were the developments in computer science that made the software and hardware available for the storage and data analysis possible. Even the legal issues of patenting have to be considered as contributing to the development of this science.

For the first time in the history of the diagnostic laboratory, molecular biology is extending the range of information available. Until this time, the laboratory has been descriptive in nature. It could measure events that were currently going on by evaluating the chemistry, hematology, or anatomical pathology. Molecular biology allows the laboratory to be predictive in nature. Now, statements can be made about events that might occur in the future. This is different from an elevated value for blood glucose, when the diagnosis of diabetes can be made. This new technology returns results that indicate the patient might be at risk for a disease. Because of the ability of this

technology to detect carriers of a mutation and predict risk for disease development, ethical considerations and genetic counseling have become an inseparable part of the laboratory procedure. Preventive medicine, therefore, will benefit from this new technology. In cases of a family history that suggests high risk for a particular disease, a lab test might indicate that there is no risk to a specific family member. If there is a significant risk, then medical care might be able to intervene at a much earlier stage. This has significant financial benefits for those that have to control the costs of healthcare.

The pharmaceutical industry will benefit greatly from the genetic engineering approach to drug production and even to synthesizing drugs specific for an individual patient. Perhaps that synthesis will take the form of modifying genes in animals so the animal will synthesize the human product, greatly reducing the need for organic synthesis.

6. OTHER FACTORS

As has been discussed, the laboratory has developed through a series of advances and setbacks. Until recently, the volume of testing has gone up dramatically every year in excess of what is predicted for diagnostic purposes. One contributing factor may be that in the mid-20th century, health insurance became available to a large portion of the working population. Under this type of program, medical care was delivered with little concern about the cost. These costs were passed along to the insurance company, which paid the charges. This put the laboratory in a revenue-generating position, and more tests translated to more income. There was no incentive to control costs; in fact, there was an incentive to do more tests, hire more staff, and buy more equipment. This led to a considerable growth in laboratory services, and although test costs were coming down, the increased volume made the total cost higher. As this new technology matures, there might be a paradoxical financial shift; that is, the individual laboratory tests might currently be more expensive to perform than more traditional assays, but the benefit from earlier intervention and genetic counseling could reduce the long-term aggregate cost of health care.

The subsequent chapters will outline a wide variety of diagnostic testing that is available, along with the benefits and medical options that are available to the patient and the health care team. This technology will carry laboratory diagnostic medicine into the 21st century.

REFERENCES

1. Rubin, L. P. *A Young Woman Taken to the Doctor by Her Family*. Mezzotint by Johann Andreas Pfeffel after an original painting by Jan Josef Horemans. *J. Hist. Med. Allied Sci.* **36(4):**488–489, 1981.
2. Haber, M. H. Pisse prophecy: a brief history of urinalysis. Clin. Lab. Med. **8(3):**415–430, 1988.
3. White, W. I. A new look at the role of urinalysis in the history of diagnostic medicine. *Clin. Chem.* **37(1):**119–125, 1991.
4. Camac, C. N. B. *Imhotep to Harvey: Backgrounds of Medical History,* Milford House, Boston, 1973.
5. Flint, A. Remarks on the use of the thermometer in diagnosis and prognosis, in *Technology and American Medical Practice 1880–1930*, Howell J. D., ed, pp. 1–14. Garland, New York, 1988.
6. Caraway, W. The scientific development of clinical chemistry to 1948. *Clin. Chem.* **19(4):**373–383, 1973.
7. Howell, J. D. *Technology in the Hospital*, The Johns Hopkins University Press. Baltimore, MD, 1995.
8. Reiser, S. J. *Medicine and the Reign of Technology*, Cambridge University Press, Cambridge, 1978.
9. Kohler, R. E. *From Medical Chemistry to Biochemistry*, Cambridge University Press, New York, 1982.
10. Chittenden, R. H. *The Development of Physiological Chemistry in the United States.* The Chemical Catalog Company, New York, 1930.
11. Meites, S. First call for clinical chemists in U.S. *Clin. Chem.* **29(10):**1852–1853, 1983.
12. Lewis, L. Leonard Skeggs—a multifaceted diamond. *Clin. Chem.* **27(8):**1465–1468, 1981.
13. Lorentz, C. P., Wieben, E. D., Tefferi, A., et al. Primer on medical genomics; Part I: History of genetics and sequencing of the human genome. *Mayo Cli. Proc.* **77,** 2002.
14. Terry, M. Spiraling through genetic history. *Adv. Med. Lab. Professionals,* May 28, 2001.
15. Morange, M. *A History of Molecular Biology*, Harvard University Press, Cambridge, MA, 1998.
16. Echols, H. *Operators and Promoters, the Story of Molecular Biology and Its Creators,* University of California Press, Berkeley, CA, 2001.
17. Baldi, P. *The Shattered Self, the End of Natural Selection*, MIT Press, Cambridge, MA, 2001.
18. Wickelgren, I. *The Gene Masters*. Times Books, Henry Holt and Co., New York, 2002.
19. Miclos, D. A. and Freyer, G. A. *DNA Science—A First Course in Recombinant DNA Technology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 4–9, 1990.
20. Blackburn, M. G. and Gait, M. J., eds. *Nucleic Acids in Chemistry and Biology,* IRL Press/Oxford University Press, New York, pp. 3–4, 1990.
21. Kay, L. E. *The Molecular Vision of Life, Caltech, The Rockefeller Foundation, and the Rise of the New Biology,* Oxford University Press, New York, 1993.

BASIC MOLECULAR BIOLOGY II

2 An Overview of Nucleic Acid Chemistry, Structure, and Function

The Foundations of Molecular Biology

WILLIAM B. COLEMAN

1. INTRODUCTION

Chemists and early biochemists determined the essential building blocks of living cells and characterized their chemical nature. Among these building blocks were nucleic acids, long-chain polymers composed of nucleotides. Nucleic acids were named based partly on their chemical properties and partly on the observation that they represent a major constituent of the cell nucleus. That nucleic acids form the chemical basis for the transmission of genetic traits was not realized until about 60 years ago *(1,2)*. Prior to that time, there was considerable disagreement among scientists as to whether genetic information was contained in and transmitted by proteins or nucleic acids. It was recognized that chromosomes contained deoxyribonucleic acid as a primary constituent, but it was not known if this DNA carried genetic information or merely served as a scaffold for some undiscovered class of proteins that carried genetic information. However, the demonstration that genetic traits could be transmitted through DNA formed the basis for numerous investigations focused on elucidation of the nature of the genetic code. During the last half-century, numerous investigators have participated in the scientific revolution leading to modern molecular biology. Of particular significance were the elucidation of the structure of DNA *(3–9)*, determination of structure–function relationships between DNA and RNA *(10,11)*, and acquisition of basic insights into the processes of DNA replication, RNA transcription, and protein synthesis *(12–19)*. Molecular pathology represents the application of the principles of basic molecular biology to the investigation of human disease processes. Our ever broadening insights into the molecular basis of disease processes continues to provide an opportunity for the clinical laboratory to develop and implement new and novel approaches for diagnosis and prognostic assessment of human disease.

2. THE CENTRAL DOGMA OF MOLECULAR BIOLOGY

Molecular biology has developed into a broad field of scientific pursuit and, at the same time, has come to represent a basic component of most other basic research sciences. This has come about through the rapid expansion of our insights into numerous basic aspects of molecular biology and the development of an understanding of the fundamental interaction among the several major processes that comprise the larger field of investigation. A theory, referred to as the "central dogma," describes the interrelationships among these major processes *(20,21)*. The central dogma defines the paradigm of molecular biology that genetic information is perpetuated as sequences of nucleic acid, but that genes function by being expressed in the form of protein molecules *(20)*. The flow of genetic information among DNA, RNA, and protein that is described by the central dogma is illustrated in Fig. 1. Individual DNA molecules serve as templates for either complementary DNA strands during the process of replication or complementary RNA molecules during the process of transcription. In turn, RNA molecules serve as blueprints for the ordering of amino acids by ribosomes during protein synthesis or translation. This simple representation of the complex interactions and interrelationships among DNA, RNA, and protein was proposed and commonly accepted shortly after the discovery of the structure of DNA. Nonetheless, this paradigm still holds more than 45 years later and continues to represent a guiding principle for molecular biologists involved in all areas of basic biological, biomedical, and genetic research.

3. CHEMICAL NATURE OF DNA

Deoxyribonucleic acid is a polymeric molecule that is composed of repeating nucleotide subunits. The order of nucleotide subunits contained in the linear sequence or primary structure of these polymers represents all of the genetic information carried by a cell. Each nucleotide is composed of (1) a phosphate group, (2) a pentose (5 carbon) sugar, and (3) a cyclic nitrogen-containing compound called a base. In DNA, the sugar moiety is 2-deoxyribose. Eukaryotic DNA is composed of four different bases: adenine, guanine, thymine, and cytosine. These bases are classified based on their chemical structure into two groups:

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

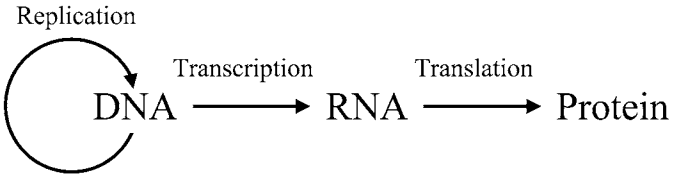


Fig. 1. The central dogma of molecular biology. The central dogma defines the paradigm of molecular biology that genetic information is perpetuated as sequences of nucleic acid, but that genes function by being expressed in the form of protein molecules. (From ref. *20*.)

Adenine and guanine are double-ring structures termed purines, and thymine and cytosine are single-ring structures termed pyrimidines (Fig. 2). Within the overall composition of DNA, the concentration of thymine is always equal to the concentration of adenine, and the concentration of cytosine is always equal to guanine *(22,23)*. Thus, the total concentration of pyrimidines always equals the total concentration of purines. These monomeric units are linked together into the polymeric structure by 3′, 5′-phosphodiester bonds (Fig. 3). Natural DNAs display widely varying sizes depending on the source. Relative molecular weights range from 1.6×10^6 Daltons for bacteriophage DNA to 1×10^{11} Daltons for a human chromosome.

4. STRUCTURE OF DNA

The structure of DNA is a double helix, composed of two polynucleotide strands that are coiled about one another in a spiral *(3,4)*. Each polynucleotide strand is held together by phosphodiester bonds linking adjacent deoxyribose moieties. The two polynucleotide strands are held together by a variety of noncovalent interactions, including lipophilic interactions between adjacent bases and hydrogen-bonding between the bases on opposite strands. The sugar-phosphate backbones of the two complementary strands are antiparallel; that is, they possess opposite chemical polarity. As one moves along the DNA double helix in one direction, the phosphodiester bonds in one strand will be oriented 5′–3′, whereas in the complementary strand, the phosphodiester bonds will be oriented 3′–5′. This configuration results in basepairs being stacked between the two chains perpendicular to the axis of the molecule. The base-pairing is always specific: Adenine is always paired to thymidine, and guanine is always paired to cytosine. This specificity results from the hydrogen-bonding capacities of the bases themselves. Adenine and thymine form two hydrogen bonds, and guanine and cytosine form three hydrogen bonds. The specificity of molecular interactions within the DNA molecule allows one to predict the sequence of nucleotides in one polynucleotide strand if the sequence of nucleotides in the complementary strand is known *(24)*. Although the hydrogen bonds themselves are relatively weak, the number of hydrogen bonds within a DNA molecule results in a very stable molecule that would never spontaneously separate under physiological conditions. There are many possibilities for hydrogen-bonding between pairs of heterocyclic bases. Most important are the hydrogen-bonded basepairs A:T and G:C that were proposed by Watson and Crick in their double-helix structure of DNA *(3,24)*. However, other forms of base-pairing have been described *(25,26)*. In addition, hydrophobic interactions between the stacked bases in the double helix lend additional stability to the DNA molecule.

Three helical forms of DNA are recognized to exist: A, B, and Z *(27)*. The B conformation is the dominate form under physiological conditions. In B DNA, the basepairs are stacked 0.34 nm apart, with 10 basepairs per turn of the right-handed double helix and a diameter of approx 2 nm. Like B DNA, the A conformer is also a right-handed helix. However, A DNA exhibits a larger diameter (2.6 nm), with 11 bases per turn of the helix, and the bases are stacked closer together in the helix (0.25 nm apart). Careful examination of space-filling models of A and B DNA conformers reveals the presence of a major groove and a minor grove *(27)*. These grooves (particularly the minor groove) contain many water molecules that interact favorably with the amino and keto groups of the bases. In these grooves, DNA-binding proteins can interact with specific DNA sequences without disrupting the base-pairing of the molecule. In contrast to the A and B conformers of DNA, Z DNA is a left-handed helix. This form of DNA has been observed primarily in synthetic double-stranded oligonucleotides, especially those with purine and pyrimidines alternating in the polynucleotide strands. In addition, high salt concentrations are required for the maintenance of the Z DNA conformer. Z DNA possesses a minor groove but no major groove, and the minor groove is sufficiently deep that it reaches the axis of the DNA helix. The natural occurrence and potential physiological significance of Z DNA in living cells has been the subject of much speculation. However, these issues with respect to Z DNA have not yet been fully resolved.

5. SEQUENCE OF THE HUMAN GENOME

The diploid genome of the typical human cell contains approx 3×10^9 basepairs of DNA that is subdivided into 23 pairs of chromosomes (22 autosomes and sex chromosomes X and Y). It has long been suggested that discerning the complete sequence of the human genome would enable the genetic causes of human disease to be investigated *(28–30)*. Practical methods for DNA sequencing appeared in the mid to late 1970s *(31–33)*, and numerous reports of DNA sequences corresponding to segments of the human genome began to appear. In the mid-1980s, a project to sequence the complete human genome was proposed, and this project began in the later years of that decade. The development of automated methods for DNA sequencing *(34,35)* made the ambitious goals of the Human Genome Project attainable *(36,37)*. Subsequently, detailed genetic and physical maps of the human genome appeared *(38–43)*, expressed sequences were identified and characterized *(44–46)*, and gene maps of the human genome were constructed *(47,48)*. Efforts by several consortiums using differing approaches *(49–51)* to large-scale sequencing of human DNA and sequence contig assembly culminated in 2001 with the publication of a draft sequence of the human genome *(52,53)*. The actual number of genes contained in the human genome is not yet known. Early estimates suggested that the human genome might contain 70,000 to 100,000 genes *(54–56)*. However, more recently, the number of genes contained in the human genome has been estimated to be approximately 30,000–40,000 *(52,53,57,58)*. Early analysis of the draft sequences of the human genome revealed considerable variability between individuals, including in excess of 1.1–1.4 million single-nucleotide polymorphisms (SNPs) distributed throughout

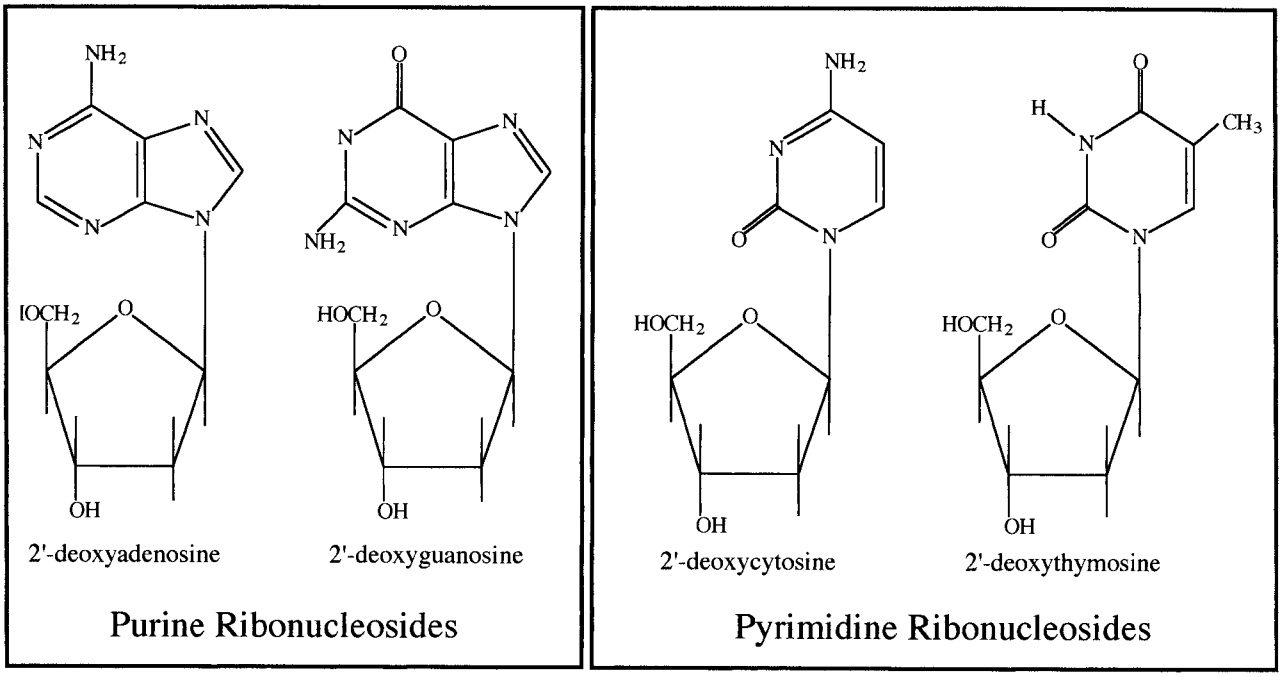


Fig. 2. The chemical structure of purine and pyrimidine deoxyribonucleosides.

the genome *(53,59)*. Refinement of the human genome sequence, identification and characterization of the genes contained, and description of the features of the genome (SNPs and other variations) continues at a rapid pace (http://www.nhgri.nih.gov/). In early 2003, it was announced that the Human Genome Project had completely sequenced 99% of the gene-containing portion of the human genome, and new plans and goals for genomics research were described *(60,61)*.

6. ORGANIZATION OF GENOMIC DNA

The human genomic DNA is packaged into discreet structural units that vary in size and genetic composition. The structural unit of DNA is the chromosome, which is a large continuous segment of DNA *(62)*. A chromosome represents a single genetically specific DNA molecule to which are attached a large number of protein molecules that are involved in the maintenance of chromosome structure and regulation of gene expression *(63)*. Genomic DNA contains both "coding" and "noncoding" sequences. Noncoding sequences contain information that does not lead to the synthesis of an active RNA molecule or protein *(54,64)*. This is not to suggest that noncoding DNA serves no function within the genome. On the contrary, noncoding DNA sequences have been suggested to function in DNA packaging, chromosome structure, chromatin organization within the nucleus, or in the regulation of gene expression *(65,66)*. A portion of the noncoding sequences represent intervening sequences that split the coding regions of structural genes. However, the majority of noncoding DNA falls into several families of repetitive DNA whose exact functions have not been entirely elucidated *(67,68)*.

Coding DNA sequences give rise to all of the transcribed RNAs of the cell, including mRNA. The organization of transcribed structural genes consists of coding regions that are interrupted by intervening noncoding regions of DNA (Fig. 4). Thus, the primary RNA transcripts contain both coding and noncoding sequences. The noncoding sequences must be removed from the primary RNA transcript during processing to produce a functional mRNA molecule appropriate for translation.

7. DNA FUNCTION

DNA serves two important functions with respect to cellular homeostasis: the storage of genetic information and the transmission of genetic information. In order to fulfill both of these functions, the DNA molecule must serve as a template. The cellular DNA provides the source of information for the synthesis of all the proteins in the cell. In this respect, DNA serves as a template for the synthesis of RNA. In cell division, DNA serves as the source of information inherited by progeny cells. In this case, DNA serves as a template for the faithful replication of the genetic information that is ultimately passed into daughter cells.

7.1. TRANSCRIPTION OF RNA Contained within the linear nucleotide sequence of the cellular DNA is the information necessary for the synthesis of all the protein constituents of a cell (Table 1). Transcription is the process in which mRNA is synthesized with a sequence complementary to the DNA of a gene to be expressed. The correct start and end points for transcription of a specific gene are identified in the DNA by a promoter sequence upstream of the gene and a termination signal downstream (Fig. 4). In the case of RNA transcription, only one strand of the DNA molecule serves as a template. This strand is referred to as the "sense" strand. Transcription of the sense strand ultimately yields a mRNA molecule that encodes the proper amino acid sequence for a specific protein.

7.2. REPLICATION OF DNA The double-stranded model of the structure of DNA strongly suggests that replication of the DNA can be achieved in a semiconservative manner *(69–72)*. In semiconservative replication, each strand of the

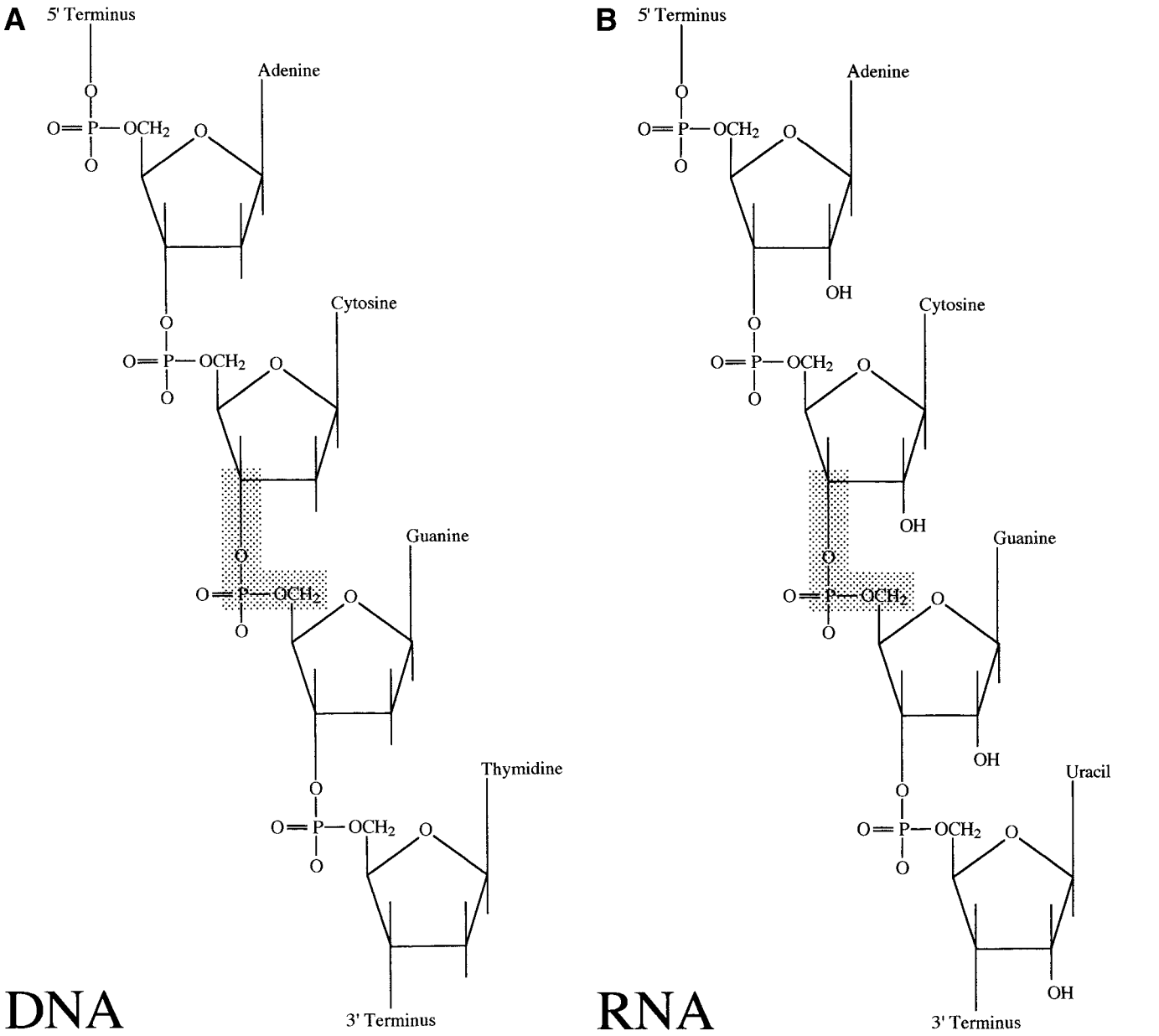


Fig. 3. The chemical structure of repeating nucleotide subunits in DNA and RNA. Each panel shows the sugar-phosphate backbone of a single polynucleotide strand of nucleic acid composed of four nucleotide subunits. The stippled area highlights a 3′–5′ phosphodiester bond.

DNA helix serves as a template for the synthesis of complementary DNA strands. The result is the formation of two complete copies of the DNA molecule, each consisting of one strand derived from the parent DNA molecule and one newly synthesized complementary strand. The utilization of the DNA strands as the template for the synthesis of new DNA ensures the faithful reproduction of the genetic material for transmission into daughter cells *(19)*.

7.3. GENETIC RECOMBINATION Genetic recombination represents one mechanism for the generation of genetic diversity through the exchange of genetic material between two homologous nucleotide sequences *(73,74)*. Such an exchange of genetic material often results in alterations of the primary structure (nucleotide sequence) of a gene and, subsequently, alteration of the primary structure of the encoded protein product. In organisms that reproduce sexually, recombination is initiated by formation of a junction between similar nucleotide sequences carried on the same chromosome from the two different parents. The junction is able to move along the DNA helix through branch migration, resulting in an exchange of the DNA strands.

7.4. DNA REPAIR Maintenance of the integrity of the informational content of the cellular DNA is absolutely required for cellular and organismal homeostasis *(75,76)*. The cellular DNA is continuously subjected to structural damage through the action of endogenous or environmental mutagens. In the absence of efficient repair mechanisms, stable mutations can be introduced into DNA during the process of replication at damaged sites within the DNA. Mammalian cells possess several distinct DNA repair mechanisms and pathways that serve to maintain DNA integrity, including enzymatic reversal repair,

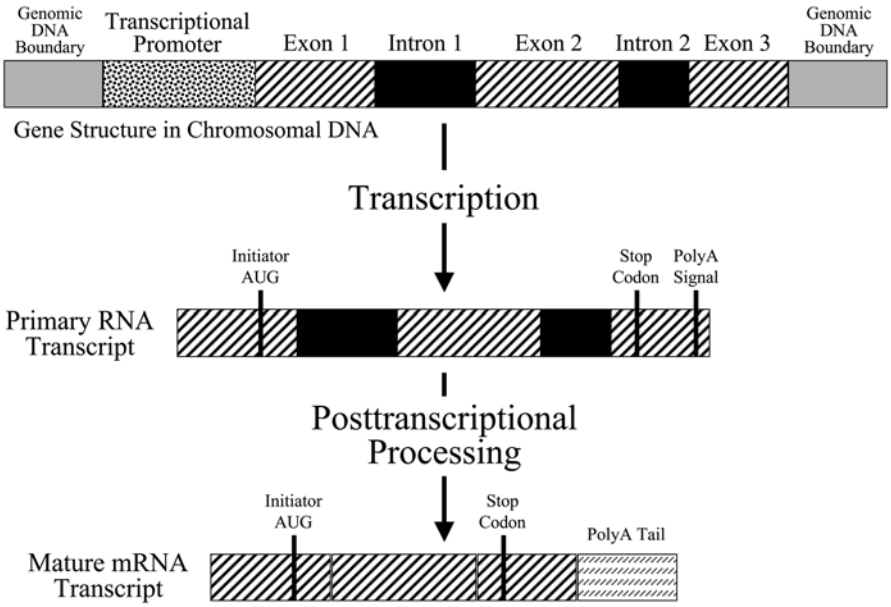


Fig. 4. Basic organization of a structural gene in DNA and biogenesis of mature mRNAs.

Table 1
The Universal Genetic Code

First position (5′ end)	*Second position*				*Third position (3′ end)*
	U	*C*	*A*	*G*	
U	Phe	Ser	Tyr	Cys	U
U	Phe	Ser	Tyr	Cys	C
U	Leu	Ser	Stop	Stop	A
U	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U
C	Leu	Pro	His	Arg	C
C	Leu	Pro	Gln	Arg	A
C	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
A	Ile	Thr	Asn	Ser	C
A	Ile	Thr	Lys	Arg	A
A	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
G	Val	Ala	Asp	Gly	C
G	Val	Ala	Glu	Gly	A
G	Val	Ala	Glu	Gly	G

nucleotide excision repair, and postreplication repair *(77,78)*. Several steps in the process of DNA repair are shared by these multiple pathways, including (1) recognition of sites of damage, (2) removal of damaged nucleotides, and (3) restoration of the normal DNA sequence. Each of these steps in DNA repair are accomplished by specific proteins and enzymes. Surveillance of the cellular DNA is a continual process involving specific aspects of the transcription and replication machinery. However, in each case where the restoration of the normal DNA sequence is accomplished through the replacement of damaged nucleotides, the undamaged DNA strand serves as a template in the repair process. This ensures the faithful reproduction of the primary structure of the DNA at the damaged site.

8. CHEMICAL NATURE OF RNA

Like DNA, RNA is composed of repeating purine and pyrimidine nucleotide subunits. However, several distinctions can be made with respect to the chemical nature of RNA and DNA. Unlike the 2′-deoxyribose sugar moiety of DNA, the sugar moiety in RNA is ribose. Like DNA, RNA usually contains adenine, guanine, and cytosine, but does not contain thymidine. In place of thymidine, RNA contains uracil. The concentration of purines and pyrimidine bases do not necessarily equal one another in RNA because of the single-stranded nature of the molecule. The monomeric units of RNA are linked together by 3′,5′-phosphodiester bonds analogous to those in DNA (Fig. 3). RNAs have molecular weights between 1×10^4 Daltons for transfer RNA (tRNA) and 1×10^7 Daltons for ribosomal RNA (rRNA).

9. STRUCTURE AND FUNCTION OF RNA

RNA exists as a long, regular, unbranched polynucleotide strand. The informational content of the RNA molecule is contained in its primary structure or nucleotide sequence. In spite of the fact that RNA exists primarily as a single-stranded molecule, significant higher-order structures are often formed in individual RNA molecules. In some cases, this higher-order structure is related to the actual function of the molecule. Three major classes of RNA are found in eukaryotic organisms: messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA). Each class differs from the others in the size, function, and general stability of the RNA molecules *(79)*. Minor classes of RNA include heterogeneous nuclear RNA (hnRNA), small nuclear RNA (snRNA), and small cytoplasmic RNA (scRNA).

9.1. STRUCTURE AND FUNCTION OF MESSENGER RNA The ability of DNA to serve directly as a template for the synthesis of protein is precluded by the observations that protein synthesis takes place in the cytoplasm, whereas almost all of the cellular DNA resides in the nucleus. Thus, genetic information contained in the DNA must be transferred to an

intermediate molecule that is translocated into the cell cytoplasm, where it directs the ordering of amino acids in protein synthesis. RNA fulfills this role as the intermediate molecule for the transport and translation of genetic information *(11)*. Messenger RNA molecules represent transcripts of structural genes that encode all of the information necessary for the synthesis of a single-type polypeptide of protein. Thus, mRNAs serve two important functions with respect to protein synthesis: (1) mRNAs deliver genetic information to the cytoplasm where protein synthesis takes place and (2) mRNAs serve as a template (or blueprint) for translation by ribosomes during protein synthesis.

Mammalian cells (among others) express "interrupted" genes; that is, genes with coding sequences are not contiguous (continuous) in the DNA, and that require a posttranscriptional modification prior to translation of protein products (Fig. 4). The majority of structural genes in the higher eukaryotic organisms are interrupted. The average gene contains 7–10 exons, spread over 10–20 kb of DNA. For instance, the *p53* tumor suppressor gene is composed of 11 exons, occupies approx 16 kb in the genomic DNA, and produces a 2-kb mRNA *(80–82)*. However, other genes are much larger. For example, the *Rb1* tumor suppressor gene occupies 200 kb in the genomic DNA, contains 27 exons, and gives rise to a 4.7-kb mRNA *(83–85)*. The primary RNA transcript exhibits the same overall structure and organization as the structural gene and is often referred to as the pre-mRNA. Removal of intronic sequences yields a mature mRNA that is considerably smaller with an average size of 1–3 kb. The process of removing the intronic sequences is called RNA splicing *(86–88)*.

The primary products of RNA transcription in the nucleus compose a special class of RNAs that are characterized by their large size and heterogeneity *(79)*. These RNA molecules are referred to as heterogeneous nuclear RNAs (hnRNAs). Heterogeneous nuclear RNAs contain both intronic and exonic sequences encoded in the template DNA of structural genes. These hnRNAs are processed in the nucleus *(87,89,90)* to give mature mRNAs that are transported into the cytoplasm, where they participate in protein synthesis. Nuclear processing of RNA involves (1) chemical modification reactions (addition of the 5′ CAP), (2) splicing reactions (removal of intronic sequences), and (3) polyadenylation [addition of the 3′ poly(A) tail]. Additional processing of some specific mRNAs occurs in the cell cytoplasm, including RNA editing reactions *(91–93)*. It has been suggested that some snRNAs function in the processing of hnRNAs *(94)*. Mature mRNAs are transported into the cytoplasm of the cell, where they participate in the translational processes of protein synthesis.

In RNA splicing, intronic sequences are specifically removed from the primary RNA transcript and the remaining exonic sequences are rejoined into one molecule. There is no extensive homology or complementarity between the two ends of an intron precluding the general possibility that intronic sequences form extensive secondary structures (such as a hairpin loop) as a preliminary step in the splicing reaction. The splice junctions represent short, well-conserved consensus sequences. The generic intron contains a GT sequence at the 5′ boundary and an AG sequence at the 3′ boundary (Fig. 5). The 5′ and 3′ splice junctions are often referred to as the splice donor and splice acceptor sites, respectively. Splice sites are generic in that they do not

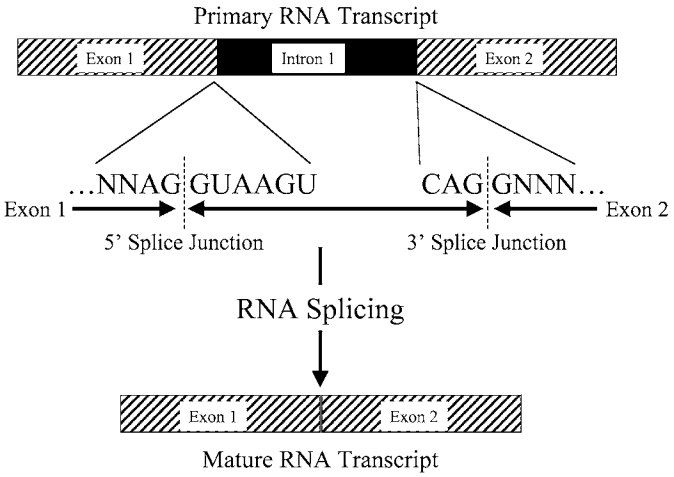


Fig. 5. Fundamental aspects of RNA splicing.

exhibit specificity for individual RNA precursors, and individual precursors do not convey specific information that is required for splicing. In principle, any splice donor site can react with any splice acceptor site. However, under normal conditions, these reactions are restricted to the donor and acceptor sites of the same intron. Analysis of molecular intermediates formed during the splicing of large precursor RNAs suggests that the introns are removed in a definitive pattern or through a preferred pathway that dictates the general order of intron removal. This suggestion might imply a mechanism in which the conformation of precursor RNA molecules limits the accessibility of splice junctions, such that as specific introns are removed, the conformation of the molecule changes and new splice sites become available. However, several other plausible mechanims exists that could account for the observed patterns of splicing in RNAs that have been examined in detail *(86,89,90)*. The RNA sequences that are required for successful splicing include (1) the 5′ splice donor and 3′ splice acceptor consensus sequences and (2) a consensus sequence known as the branch site. The branch site is located approx 20–40 bases upstream of the 3′-terminus of the intronic sequence and conforms to the following consensus sequence: Py-N-Py-Py-Pu-A-Py. The role of the branch site is to identify the nearest splice acceptor site for connection to the splice donor site. The first step in splicing involves a cleavage of the RNA molecule at the 5′ end of the intron. The resulting intron–exon molecule forms a structure known as a lariat, through the formation of a 5′–2′ bond between the G residue located at the 5′ end of the intron and the adenine residue of the branch site. The second step involves cleavage of the RNA molecule at the 3′ splice site This cleavage releases the intron in lariat form, which is subsequently degraded.

The 5′-termini of all eukaryotic mRNAs are modified posttracriptionally, through enzymatic reactions occuring in the nucleus and in the cytoplasm of the cell. Initially, a guanine residue is added to the 5′-termini of primary mRNA transcripts through the action of an enzyme called guanylyl transferase. This reaction occurs in the nucleus soon after the intitiation of transcription. This guanine residue is linked to the first coded

nucleoside triphosphate through a 5′–5′ triphosphate linkage, rather than a 3′–5′ phosphodiester bond. Thus, this guanine residue occurs in the structure of the mRNA in reverse orientation from all the other nucleotides. The modified 5′-terminus is referred to as a "CAP" and is the site of additional modification reactions involving the addition of methyl groups. These additional modifications are catalyzed by enzymes located in the cytoplasm of the cell. The first methylation occurs in all mRNAs and consists of the addition of a methyl group to the 7-position of the terminal guanine through the action of guanine-7-methyltransferase. Additional methylation reactions can occur involving the additional bases at the 5′-terminus of the mRNA transcript, and less frequently, internal methylation of bases within an mRNA molecule takes place.

The poly(A) tail possessed by most eukaryotic mRNAs is not encoded in the DNA; rather, it is added to the RNA in the nucleus after transcription of the structural gene is complete. The addition of the poly(A) tail is catalyzed by the enzyme poly(A) polymerase, which adds approx 200 adenosine residues to the free 3′-OH terminus of the primary RNA transcript. The precise function of the poly(A) tail is unknown, but has been speculated to be involved in mRNA stability or in control of mRNA utilization *(95)*. Although the removal of the poly(A) tail does precede degradation of certain mRNAs, a systematic correlation between mRNA stability or survival, and the length or presence of the poly(A) tail has not been established. Removal of the poly(A) tail can inhibit the initiation of translation in vitro, suggesting a potential role for this structure in the control of mRNA translation. However, it is not clear whether this effect is related to a direct influence of poly(A) structures on the initiation reaction or the result of some indirect cause.

Some other forms of posttranscriptional RNA processing occur with respect to a small subset of eukaryotic mRNAs. The process of RNA editing involves a posttranscriptional alteration of the informational content of a specific mRNA *(91)*. The editing of RNA is revealed when the linear sequence of the mRNA molecule differs from the coding sequence carried in the DNA. In mammalian cells, there are examples where substitution of a single base occurs in the mRNA, resulting in an alteration of an amino acid in the final protein product. Because no known template source mediates the RNA editing reaction, the most likely mechanism for mRNA editing would involve a specific enzyme that can recognize the sequence or secondary structure of the specific target mRNA and catalyze the specific base substitution. However, there are examples of RNA editing in some lower eukaryotes that utilize "guide RNA," which directs the RNA editing reaction *(96)*. The final result of RNA editing is the generation protein products representing more than one polypeptide sequence from a single coding gene. The different protein products might possess different biological activities, suggesting that RNA editing might represent a mechanism for controlling the functional expression of genes through a posttranscriptional process that does not impact the normal mechanisms for controlling levels of gene expression.

9.2. STRUCTURE AND FUNCTION OF TRANSFER RNA Transfer RNAs are small molecules consisting of approx 75–80 nucleotides. Like mRNAs, tRNA is generated through nuclear processing of precursor RNA transcripts. The structure of the tRNA molecule reflects its function as an adapter between the mRNA and amino acids during protein synthesis. Specific tRNAs correspond to each of the amino acids utilized in protein synthesis in any particular cell type. Although the specific tRNAs differ from each other with respect to their actual nucleotide sequence, tRNAs as a class of RNA molecules share several common structural features. Each tRNA contains information in the primary structure or nucleotide sequence that dictates the higher-order structure of the molecule. The secondary structure of the tRNA resembles a cloverleaf *(97)*. The folding of the cloverleaf structure is maintained through intrastrand sequence complementarity and base-pairing interactions between nucleotides. In addition, each tRNA contains an ACC sequence at the 3′-terminus and an anticodon loop *(97)*. Amino acids are attached to their specific tRNA through an ester bond to the 3′-hydroxyl group of the terminal adenine of the ACC sequence. The anticodon loop recognizes the triplet codon of the template mRNA during the process of translation. With the exception of the codons encoding methionine and tryptophan, there are at least two possible codons for each amino acid (Table 1). Nonetheless, each amino acid has only one corresponding tRNA. Thus, the hydrogen-bonding between nucleotides of the codon and anticodon often involve "wobble" pairing *(98)*. This form of base-pairing allows mismatches in the third base of a codon triplet. In the overall structure–function relationship in tRNA, the nucleotide sequence of the anticodon loop dictates which amino acid will be attached to the ACC sequence of the tRNA.

Transfer RNAs serve as adapters between the mRNA template and the amino acids of growing polypeptide chains during the process of protein translation *(97)*; that is, the tRNA serves to ferry the appropriate amino acid into the active site of the ribosome, where it becomes incorporated into the growing polypeptide being synthesized. Amino acids are coupled to their specific tRNA through the action of enzymes called aminoacyl tRNA synthetases. The specificity of the "charging" reaction is critical to the integrity of the translation process because the incorporation of amino acids at the level of the ribosome depends wholly on the sequence of the anticodon portion of the tRNA molecule. The charged tRNA the mRNA through the transient hybridization of the codon and anti-codon RNA sequences in the ribosome complex as it moves along the mRNA. The entry of the charged tRNA into the active site of the ribosome brings its associated amino acid into juxtaposition with the nascent polypeptide, facilitating the formation of a peptide bond. In this manner, the tRNA provides a link between the genetic information contained in the mRNA and the linear sequence of amino acids represented in the resulting polypeptide product.

9.3. STRUCTURE AND FUNCTION OF RIBOSOMAL RNA The ribosome is a nucleoprotein that serves as the primary component of a cell's protein synthesis machinery. The ribosome is a complex structure consisting of two subunit particle types (60S and 40S). The overall composition of the fully assembled ribosome includes at least 4 distinct rRNA molecules and nearly 100 specific protein subunits. The major rRNAs in mammalian cells were named for their molecular size as determined by their sedimentation rates. Three of these rRNAs (5S,

Table 2
Secondary and Tertiary Structural Motifs in RNA

Structure	*Description*	*Functions*
Hairpin loops	Single-stranded loop that bridges one end of a double-stranded stem	Component of more complex RNA structures; May serve as a nucleation site for RNA folding, or recognition sites for protein–RNA interaction
Internal loops	Interruptions in double-stranded RNA caused by the presence of nucleotides on both strands that cannot participate in Watson–Crick base-pairing	Protein-binding sites and ribozyme cleavage sites
Bulges	Double-stranded RNA molecules with unpaired nucleotides on only one strand	Contribute to the formation of more complex RNA structures; recognition sites for protein–RNA interactions.
Nucleotide triples	Triple helical structures that form through hydrogen bonding between nucleotides of a single-stranded RNA molecule and nucleotides within a double-stranded RNA molecule; hydrogen bonds can involve nucleotide bases, sugars, or phosphate groups	Orient regions of secondary structure in large RNA molecules and stabilize three-dimensional RNA structures.
Pseudoknots	Results from base-pairing between nucleotide sequences within an RNA loop structure and a complementary nucleotide sequence outside the RNA loop	RNA self-splicing, autoregulation of translational processes, and ribosome frameshifting

5.8S, and 28S) are components of 60S ribosomal particle. The smaller 40S ribosomal particle contains a single 18S rRNA. The 5.8S, 18S, and 28S rRNAs are the products of the processing of a single 45S precursor RNA molecule. The 5S rRNA is independently transcribed and processed. The rRNAs assemble with ribosomal protein subunits in the nucleus. The precise role of rRNA in the function of the ribosome is not completely understood *(99)*. However, it is recognized that interactions between the rRNAs of the ribosomal subunits might be important in the overall structure of the functioning ribosomal particle. In addition, rRNA sequences can interact with ribosome-binding sequences of mRNA during the initiation of translation. Likewise, it is likely that rRNAs bind to invariant tRNA sequences when these molecules enter the active site of the ribosome *(99)*.

9.4. SPECIAL RNA STRUCTURES Higher-order RNA structures exhibit hydrogen-bonding between A:U and G:C basepairs. Several specific higher-order RNA structures have been recognized (Table 2) and characterized in detail *(100)*. Hairpin loops consist of a double-stranded stem and a single-stranded loop that bridges one end of the stem. These structures are essential components of more complex RNA structures and probably serve as nucleation sites for RNA folding. Loops can also function as recognition sites for protein–RNA interactions. Internal loops represent interruptions in double-stranded RNA caused by the presence of nucleotides on both strands that cannot participate in Watson–Crick base-pairing. Several important functions are associated with internal loops, including protein-binding sites and ribozyme cleavage sites. In many cases, internal loops have been shown to represent highly ordered structures maintained by the formation of non-Watson–Crick basepairs. Bulges are structural motifs contained within double-stranded RNA molecules with unpaired nucleotides on only one strand. RNA bulges contribute to the formation of more complex, higher-order RNA structures and can also serve as recognition sites for protein–RNA interaction. Nucleotide triples occur when single-stranded RNA sequences form hydrogen bonds with nucleotides that are already base-paired. These interactions serve to stabilize three-dimensional RNA structures and to orient regions of RNA secondary structure in large RNA molecules. Pseudoknots are tertiary structural elements that result from base-pairing between nucleotide sequences contained within a loop structure and sequences outside the loop structure. These are important in RNA self-splicing, translational autoregulation, and ribosomal frameshifting.

10. DNA DAMAGE, MUTAGENESIS, AND THE CONSEQUENCES OF MUTATION

DNA damage can result from spontaneous alteration of the DNA molecule or from the interaction of numerous chemical and physical agents with the structural DNA molecule *(101)*. Spontaneous lesions can occur during normal cellular processes, such as DNA replication, DNA repair, or gene rearrangement *(78)*, or through chemical alteration of the DNA molecule itself as a result of hydrolysis, oxidation, or methylation *(102,103)*. In most cases, DNA lesions create nucleotide mismatches that lead to point mutations. Nucleotide mismatches can result from the formation of apurinic or apyrimidinic sites following depurination or depyrimidination reactions *(103)*, nucleotide conversions involving deamination reactions *(78)*, or, in rare instances, from the presence of a tautomeric form of an individual nucleotide in replicating DNA. Deamination reactions can result in the conversion of cytosine to uracil, adenine to hypoxanthine, and guanine to xanthine *(78)*. However, the most common nucleotide deamination reaction involves methylated cytosines, which can replace cytosine in the linear sequence of a DNA molecule in the form of 5-methylcytosine. The 5-methylcytosine residues are always located next to guanine residues on the same chain, a motif

referred to as a CpG island. The deamination of 5′-methylcytosine results in the formation of thymine. This particular deamination reaction accounts for a large percentage of spontaneous mutations in human disease *(104–106)*. Interaction of DNA with physical agents, such as ionizing radiation, can lead to single- or double-strand breaks as a result of scission of phosphodiester bonds on one or both polynucleotide strands of the DNA molecule *(78)*. Ultraviolet (UV) light can produce different forms of photoproducts, including pyrimidine dimers between adjacent pyrimidine bases on the same DNA strand. Other minor forms of DNA damage caused by UV light include strand breaks and crosslinks *(78)*. Nucleotide base modifications can result from exposure of the DNA to various chemical agents, including N-nitroso compounds and polycyclic aromatic hydrocarbons *(78)*. DNA damage can also be caused by chemicals that intercalate the DNA molecule and/or crosslink the DNA strands *(78)*. Bifunctional alkylating agents can cause both intrastrand and interstrand crosslinks in the DNA molecule.

The various forms of spontaneous and induced DNA damage can give rise to a plethora of different types of molecular mutation *(107)*. These various types of mutation include both gross alteration of chromosomes and more subtle alterations to specific gene sequences in otherwise normal chromosomes. Gross chromosomal aberrations include (1) large deletions, (2) additions (reflecting amplification of DNA sequences), and (3) translocations (reciprocal and nonreciprocal). All of these forms of chromosomal abnormality can be distinguished through standard karyotype analyses of G-banded chromosomes (Fig. 6). The major consequence of chromosomal deletion is the loss of specific genes that are located in the deleted chromosomal region, resulting in changes in the copy number of the affected genes. The deletion of certain classes of genes such as tumor suppressor genes or genes encoding the proteins involved in DNA repair can predispose cells to neoplastic transformation *(108,109)*. Likewise, amplification of chromosomal regions results in an increase in gene copy numbers, which can lead to the same type of circumstance if the affected region contains genes for dominant proto-oncogenes or other positive mediators of cell cycle progression and proliferation *(108–110)*. The direct result of chromosomal translocation is the movement of some segment of DNA from its natural location into a new location within the genome, which can result in altered expression of the genes that are contained within the translocated region. If the chromosomal breakpoints utilized in a translocation are located within structural genes, then new hybrid genes can be generated.

The most common forms of mutation involve single-nucleotide alterations, small deletions, or small insertions into specific gene sequences. These microscopic alterations very often can only be detected through DNA sequencing. Single-nucleotide alterations that involve a change in the normal coding sequence of the gene are referred to as point mutations. The consequence of most point mutations is an alteration in the amino acid sequence of the encoded protein. However, some point mutations are "silent" and do not affect the structure of the gene product (Table 3). Silent mutations are possible because most amino acids can be encoded by more than one triplet codon (Table 1). Point mutations fall into two classes:

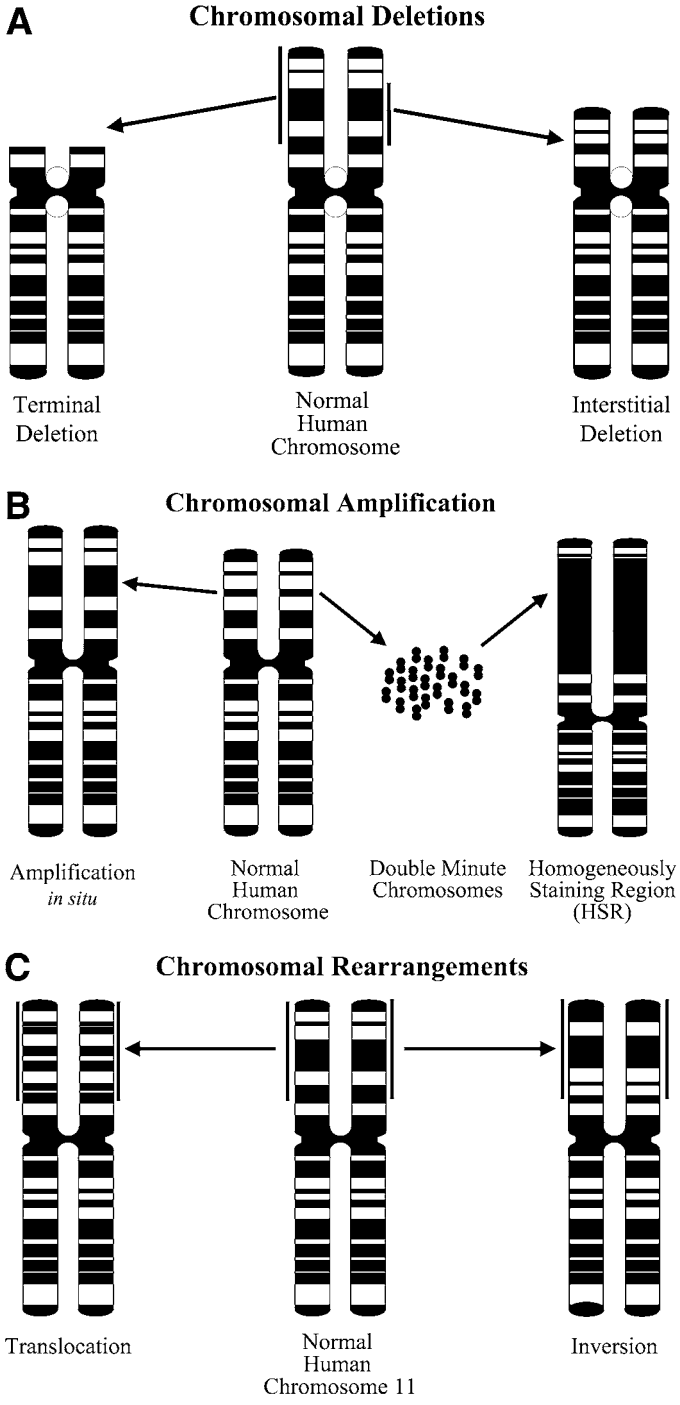


Fig. 6. Forms of gross chromosomal aberrations. Three examples of chromosomal aberrations are demonstrated using a standard G-band ideogram of human chromosome 11. The asterisks denote the chromosomal band that has expanded as a result of DNA amplification. The stippled area highlights a region from another chromosome that has been translocated to the long arm of the starting chromosome, displacing the normal qter region without altering the overall size of the chromosome (note the differences in G-banding in this region).

missense mutations and nonsense mutations. Missense mutations involve nucleotide base substitutions that alter the translation of the affected codon triplet. In contrast, nonsense mutations involve nucleotide base substitutions that modify a triplet codon that normally encodes for an amino acid into a

Table 3
Forms and Consequences of Molecular Mutation

Normal coding sequence and amino acid translation									
. . . Phe	Phe	Glu	Pro	Gly	Ser	Asn	Val	Tyr	. . .
. . . UUC	UUU	GAA	CCG	GGA	AGC	AAU	GUC	UAC	A. . .
Missense point mutation resulting in amino acid change									
. . . Phe	Phe	Glu	Pro	**Val**	Ser	Asn	Val	Tyr	. . .
. . . UUC	UUU	GAA	CCG	G**C**A	AGC	AAU	GUC	UAC	A. . .
Missense point mutation without amino acid change (silent mutation)									
. . . Phe	Phe	Glu	Pro	Gly	Ser	Asn	Val	Tyr	. . .
. . . UUC	UUU	GAA	CCG	GG**C**	AGC	AAU	GUC	UAC	A. . .
Frameshift mutation resulting from a single base insertion									
. . . Phe	Phe	Glu	Pro	**Arg**	**Lys**	**Gln**	**Cys**	**Leu**	. . .
. . . UUC	UUU	GAA	CCG	**A**GG	AAG	CAA	UGU	CUA	CA. . .
Frameshift mutation resulting from a single base deletion									
. . . Phe	Phe	Glu	Pro	**Glu**	**Asp**	**Met**	**Ser**	**Asn**	. . .
. . . UUC	UUU	GAA	CCG	GAA	GCA	AUG	UCU	ACA	. . .
Nonsense mutation resulting in a premature stop codon									
. . . Phe	Phe	Glu	Pro	**Stop**					. . .
. . . UUC	UUU	GAA	CCG	UGA	AGC	AAU	GUC	UAC	. . .

translational stop codon. This results in the premature termination of translation and the production of a truncated protein product. Small deletions and insertions can usually be classified as frameshift mutations because the deletion or insertion of a single nucleotide (for instance) alters the reading frame of the gene on the 3′ side of the affected site. This results in the synthesis of a polypeptide product that might bear no resemblance to the normal gene product (Table 3). In addition, small insertions or deletions can result in the premature termination of translation resulting from the presence of a stop codon in the new reading frame of the mutated gene. Deletions or insertions that occur involving multiples of three nucleotides will not result in a frameshift mutation, but will alter the resulting polypeptide gene product, which will exhibit either loss of specific amino acids or the presence of additional amino acids within its primary structure. These types of alteration can also lead to a loss of protein function.

REFERENCES

1. Avery, O. T., MacLeod, C. M., and McCarty, M. Studies on the chemical nature of the substance inducing transformation of Pneumococcus types. Induction of transformation by a desoxyribonucleic acid fraction isolated from Pneumococcus Type III. *J. Exp. Med.* **79:**137–158, 1944.
2. McCarty, M. and Avery, O. T. Studies of the chemical nature of the substance inducing transformation of pneumococcal types II. Effect of desoxyribonulcease on the biological activity of the transforming substance. *J. Exp. Med.* **83:**89–96, 1946.
3. Watson, J. D. and Crick F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171:**737–738, 1953.
4. Wilkins, M. H., Stokes, A. R., and Wilson, H. R. Molecular structure of deoxypentose nucleic acids. *Nature* **171:**738–740, 1953.
5. Franklin, R. E. and Gosling, R. G. Molecular structure of nucleic acids. Molecular configuration in sodium thymonucleate. 1953. *Ann. NY Acad. Sci.* **758:**16–17, 1995.
6. Watson, J. D. and Crick, F. H. The structure of DNA. *Cold Spring Harbor Symp. Quant. Biol.* **18:**123–131, 1953.
7. Franklin, R. E. and Gosling, R. G. Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature* **172:**156–157, 1953.
8. Jacobson, B. Hydration structure of deoxyribonucleic acid and its physicochemical properties. *Nature* **172:**666–667, 1953.
9. Wilkins, M. H., Seeds, W. E., Stokes, A. R., and Wilson, H. R. Helical structure of crystalline deoxypentose nucleic acid. *Nature* **172:**759–762, 1953.
10. Watson, J. D. and Crick, F. H. Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171:**964–967, 1953.
11. Brenner, S., Jacob, F., and Meselson, M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* **190:**576–581, 1961.
12. Dounce, A. L. Duplicating mechanism for peptide chain and nucleic acid synthesis. *Enzymologia* **15:**251–258, 1952.
13. Lehman, I. R., Bessman, M. J., Simms, E. S., and Kornberg, A. Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from *Escherichia coli*. *J. Biol. Chem.* **233:**163–170, 1958.
14. Kornberg, A. Biologic synthesis of deoxyribonucleic acid. *Science* **131:**1503–1508, 1960.
15. Nirenberg, M. W. and Matthaei, J. H. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. USA* **47:**1588–1602, 1961.
16. Matthaei, H. and Nirenberg, M. W. The dependence of cell-free protein synthesis in *E. coli* upon RNA prepared from ribosomes. *Biochem. Biophys. Res. Commun.* **4:**404–408, 1961.
17. Kornberg, R. D. and Lorch, Y. Chromatin structure and transcription. *Annu. Rev. Cell Biol.* **8:**563–587, 1992.
18. Kozak, M. Regulation of translation in eukaryotic systems. *Annu. Rev. Cell Biol.* **8:**197–225, 1992.
19. Coverley, D. and Laskey, R. A. Regulation of eukaryotic DNA replication. *Annu. Rev. Biochem.* **63:**745–776, 1994.
20. Crick, F. H. C. On protein synthesis. *Symp. Soc. Exp. Biol.* **12:**548–555, 1958.
21. Crick, F. Central dogma of molecular biology. *Nature* **227:**561–563, 1970.
22. Chargaff, E., Vischer, E., Doniger, R., Green, C., and Misani, F. The composition of the desoxypentose nucleic acids of thymus and spleen. *J. Biol. Chem.* **177:**405–416, 1949.

23. Chargaff, E. Structure and function of nucleic acids as cell constituents. *Fed. Proc.* **10:**654–659, 1951.
24. Crick, F. H. C. and Watson, J. D. The complementary structure of deoxyribonucleic acid. *Proc. R. Soc. A* **223:**80–96, 1954.
25. Hunter, W. N., Brown, T., Anand, N. N., and Kennard. O. Structure of an adenine–cytosine base pair in DNA and its implications for mismatch repair. *Nature* **320:**552–555, 1986.
26. Liu, K., Miles, H. T., Frazier, J., and Sasisekharan, V. A novel DNA duplex. A parallel-stranded DNA helix with Hoogsteen base pairing. *Biochemistry* **32:**11,802–11,809, 1983.
27. Dickerson, R. E., Drew, H. R., Conner, B. N., Wing, R. M., Fratini, A. V., and Kopka, M. L. The anatomy of A-, B-, and Z-DNA. *Science* **216:**475–485, 1982.
28. Collins, F. S. The human genome project and the future of medicine. *Ann. NY Acad. Sci.* **882:**42–55, 1999; discussion 56–65.
29. Futreal, P. A., Kasprzyk, A., Birney, E., Mullikin, J. C., Wooster, R., and Stratton, M. R. Cancer and genomics. *Nature* **409:**850–852, 2001.
30. Jimenez-Sanchez, G., Childs, B., and Valle, D. Human disease genes. *Nature* **409:**853–855, 2001.
31. Sanger, F. and Coulson, A. R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* **94:**441–448, 1975.
32. Sanger, F., Nicklen, S., and Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74:**5463–5467, 1977.
33. Maxam, A. M. and Gilbert, W. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* **74:**560–564, 1977.
34. Smith, L. M., Sanders, J. Z., Kaiser, R. J., et al. Fluorescence detection in automated DNA sequence analysis. *Nature* **321:**674–679, 1986.
35. Ansorge, W., Sproat, B. S., Stegemann, J., and Schwager, C. A non-radioactive automated method for DNA sequence determination. *J. Biochem. Biophys. Methods* **13:**315–323, 1986.
36. Collins, F. and Galas, D. A new five-year plan for the U.S. Human Genome Project. *Science* **262:**43–46, 1993.
37. Collins, F. S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., and Walters, L. New goals for the U.S. Human Genome Project: 1998–2003. *Science* **282:**682–689, 1998.
38. Hudson, T. J., Stein, L. D., Gerety, S. S., et al. An STS-based map of the human genome. *Science* **270:**1945–1954, 1995.
39. Berry, R., Stevens, T. J., Walter, N. A., et al. Gene-based sequence-tagged-sites (STSs) as the basis for a human gene map. *Nat. Genet.* **10:**415–423, 1995.
40. Gyapay, G., Schmitt, K., Fizames, C., et al. A radiation hybrid map of the human genome. *Hum. Mol. Genet.* **5:**339–346, 1996.
41. Cheung, V. G., Nowak, N., Jang, W., et al. Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature* **409:**953–958, 2001.
42. Olivier, M., Aggarwal, A., Allen, J., et al. A high-resolution radiation hybrid map of the human genome draft sequence. *Science* **291:**1298–1302, 2001.
43. McPherson, J. D., Marra, M., Hillier, L., et al. A physical map of the human genome. *Nature* **409:**934–941, 2001.
44. Adams, M. D., Dubnick, M., Kerlavage, A. R., et al. Sequence identification of 2,375 human brain genes. *Nature* **355:**632–634, 1992.
45. Adams, M. D., Soares, M. B., Kerlavage, A. R., Fields, C., and Venter, J. C. Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. *Nat. Genet.* **4:**373–380, 1993.
46. Adams, M. D., Kerlavage, A. R., Fields, C., and Venter, J. C. 3,400 new expressed sequence tags identify diversity of transcripts in human brain. *Nat. Genet.* **4:**256–267, 1993.
47. Schuler, G. D., Boguski, M. S., Stewart, E. A., et al. A gene map of the human genome. *Science* **274:**540–546, 1996.
48. Caron, H., van Schaik, B., van der Mee, M., et al. The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science* **291:**1289–1292, 2001.
49. Venter, J. C., Adams, M. D., Martin-Gallardo, A., McCombie, W. R., and Fields, C. Genome sequence analysis: scientific objectives and practical strategies. *Trends Biotechnol.* **10:**8–11, 1992.
50. Venter, J. C., Smith, H. O., and Hood, L. A new strategy for genome sequencing. *Nature* **381:**364–366, 1996.
51. Venter, J. C., Adams, M. D., Sutton, G. G., et al. Shotgun sequencing of the human genome. *Science* **280:**1540–1542, 1998.
52. Lander, E. S., Linton, L. M., Birren, B., et al. Initial sequencing and analysis of the human genome. *Nature* **409:**860–921, 2001.
53. Venter, J. C., Adams, M. D., Myers, E. W., et al. The sequence of the human genome. *Science* **291:**1304–1351, 2001.
54. Nowak, R. Mining treasures from 'junk DNA.' *Science* **263:**608–610, 1994.
55. Antequera, F. and Bird, A. Predicting the total number of human genes. *Nat. Genet.* **8:**114, 1994.
56. Fields, C., Adams, M. D., White, O., and Venter, J. C. How many genes in the human genome? *Nat. Genet.* **7:**345–346, 1994.
57. Ewing, B. and Green, P. Analysis of expressed sequence tags indicates 35,000 human genes. *Nat. Genet.* **25:**232–234, 2000.
58. Roest Crollius, H., Jaillon, O., Bernot, A. Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence. *Nat. Genet.* **25:**235–238, 2000.
59. Sachidanandam, R., Weissman, D., Schmidt, S. C., et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409:**928–933, 2001.
60. Collins, F. S., Green, E. D., Guttmacher, A. E., and Guyer, M. S. A vision for the future of genomics research. *Nature* **422:**835–847, 2003.
61. Collins, F. S., Morgan, M. and Patrinos, A. The Human Genome Project: lessons from large-scale biology. *Science* **300:**286–290, 2003.
62. Kavenoff, R., Klotz, L. C., and Zimm, B. H. One the nature of chromosome-sized DNA molecules. *Cold Spring Harbor Symp. Quant. Biol.* **38:**1–8, 1974.
63. Spector, D. L. The dynamics of chromosome organization and gene regulation. *Annu. Rev. Biochem.* **72:**573–608, 2003.
64. Mattick, J. S. Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep.* **2:**986–991, 2001.
65. Small, D., Nelkin, B., and Vogelstein, B. Nonrandom distribution of repeated DNA sequences with respect to supercoiled loops and the nuclear matrix. *Proc. Natl. Acad. Sci. USA* **79:**5911–5915, 1982.
66. Tsongalis, G. J., Coleman, W. B., Smith, G. J., and Kaufman, D. G. Partial characterization of nuclear matrix attachment regions from human fibroblast DNA using Alu-polymerase chain reaction. *Cancer Res.* **52:**3807–3810, 1992.
67. Jelinek, W. R. and Schmid, C. W. Repetitive sequences in eukaryotic DNA and their expression. *Annu. Rev. Biochem.* **51:**813–844, 1982.
68. Schmid, C. W. and Jelinek, W. R. The Alu family of dispersed repetitive sequences. *Science* **216:**1065–1070, 1958.
69. Meselson, M. and Stahl, F. W. The replication of DNA. *Cold Spring Harbor Symp. Quant. Biol.* **23:**9–12, 1958.
70. Meselson, M. and Stahl, F. W. The replication of DNA in *Escherichia coli. Proc. Natl. Acad. Sci. USA* **44:**671–682, 1958.
71. Franceschini, P. Semiconservative DNA duplication in human chromosomes treated with BUDR and stained with acridine orange. *Exp. Cell Res.* **89:**420–421, 1974.
72. Rude, J. M. and Friedberg, E. C. Semi-conservative deoxyribonucleic acid synthesis in unirradiated and ultraviolet-irradiated xeroderma pigmentosum and normal human skin fibroblasts. *Mutat. Res.* **42:**433–442, 1977.
73. West, S. C. Enzymes and molecular mechanisms of genetic recombination. *Annu. Rev. Biochem.* **61:**603–640, 1992.
74. Alberts, B. DNA replication and recombination. *Nature* **421:**431–435, 2003.
75. Friedberg, E. C. Biological responses to DNA damage: a perspective in the new millennium. *Cold Spring Harb. Symp. Quant. Biol.* **65:**593–602, 2000.
76. Friedberg, E. C. DNA damage and repair. *Nature* **421:**436–440, 2003.
77. Sancar, A. and Sancar G. B. (1988) DNA repair enzymes. *Annu. Rev. Biochem,* **57:**29–67.

78. Friedberg, E. C., Walker, G. C., and Siede, W. *DNA Repair and Mutagenesis*, ASM, Washington, DC, pp. xvii, 1995.
79. Lewin, B. Units of transcription and translation: sequence components of heterogeneous nuclear RNA and messenger RNA. *Cell* **4:**77–93, 1975.
80. Oren, M. The p53 cellular tumor antigen: gene structure, expression and protein properties. *Biochim. Biophys. Acta* **823:**67–78, 1985.
81. Hollstein, M., Hergenhahn, M., Yang, Q., Bartsch, H., Wang, Z. Q., and Hainaut, P. New approaches to understanding p53 gene tumor mutation spectra. *Mutat. Res.* **431:**199–209, 1999.
82. Bennett, W. P., Hussain, S. P., Vahakangas, K. H., Khan, M. A., Shields, P. G., and Harris, C. C. Molecular epidemiology of human cancer risk: gene–environment interactions and p53 mutation spectrum in human lung cancer. *J. Pathol.* **187:**8–18, 1999.
83. Bookstein, R., Lee, E. Y., To, H., et al. Human retinoblastoma susceptibility gene: genomic organization and analysis of heterozygous intragenic deletion mutants. *Proc. Natl. Acad. Sci. USA* **85:**2210–2214, 1988.
84. Lee, W. H., Bookstein, R., Hong, F., Young, L. J., Shew, J. Y., and Lee, E. Y. Human retinoblastoma susceptibility gene: cloning, identification, and sequence. *Science* **235:**1394–1399, 1987.
85. Hong, F. D., Huang, H. J., To, H., et al. Structure of the human retinoblastoma gene. *Proc. Natl. Acad. Sci. USA* **86:**5502–5506, 1989.
86. Cech, T. R. Self-splicing of group I introns. *Annu. Rev. Biochem.* **59:**543–568, 1990.
87. Jurica, M. S. and Moore, M. J. Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell* **12:**5–14, 2003.
88. Black, D. L. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72:**291–336, 2003.
89. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S., and Sharp, P. A. Splicing of messenger RNA precursors. *Annu. Rev. Biochem.* **55:**1119–1150, 1986.
90. Sharp, P. A. Splicing of messenger RNA precursors. *Science* **235:**766–771, 1987.
91. Powell, L. M., Wallis, S. C., Pease, R. J., Edwards, Y. H., Knott, T. J., and Scott, J. A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell* **50:**831–840, 1987.
92. Sollner-Webb, B. RNA editing. *Curr. Opin. Cell Biol.* **3:**1056–1061, 1991.
93. Landweber, L. F. and Gilbert, W. RNA editing as a source of genetic variation. *Nature* **363:**179–182, 1993.
94. Mattaj, I. W., Tollervey, D., and Seraphin, B. Small nuclear RNAs in messenger RNA and ribosomal RNA processing. *FASEB J.* **7:**47–53, 1993.
95. Darnell, J. E., Philipson, L., Wall, R., and Adesnik, M. Polyadenylic acid sequences: role in conversion of nuclear RNA into messenger RNA. *Science* **174:**507–510, 1971.
96. Blum, B., Bakalara, N., and Simpson, L. A model for RNA editing in kinetoplastid mitochondria: "guide" RNA molecules transcribed from maxicircle DNA provide the edited information. *Cell* **60:**189–198, 1990.
97. Schimmel, P. and Ribas de Pouplana, L. Transfer RNA: from minihelix to genetic code. *Cell* **81:**983–986, 1995.
98. Crick, F. H. Codon–anticodon pairing: the wobble hypothesis. *J. Mol. Biol.* **19:**548–555, 1966.
99. Noller, H. F. Structure of ribosomal RNA. *Annu. Rev. Biochem.* **53:**119–162, 1984.
100. Shen, L. X., Cai, Z., and Tinoco, I., Jr. RNA structure at high resolution. *FASEB J.* **9:**1023–1033, 1995.
101. Drake, J. W. and Baltz, R. H. The biochemistry of mutagenesis. *Annu. Rev. Biochem.* **45:**11–37, 1976.
102. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362:**709–715, 1993.
103. Ames, B. N., Shigenaga, M. K., and Gold, L. S. DNA lesions, inducible DNA repair, and cell division: three key factors in mutagenesis and carcinogenesis. *Environ. Health Perspect.* **101(Suppl 5):**35–44, 1993.
104. Cooper, D. N. and Youssoufian, H. The CpG dinucleotide and human genetic disease. *Hum. Genet.* **78:**151–155, 1988.
105. Rideout, W. M., 3rd, Coetzee, G. A., Olumi, A. F., and Jones, P. A. 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes. *Science* **249:**1288–1290, 1990.
106. Jones, P. A., Buckley, J. D., Henderson, B. E., Ross, R. K., and Pike, M. C. From gene to carcinogen: a rapidly evolving field in molecular epidemiology. *Cancer Res.* **51:**3617–3620, 1991.
107. Bishop, J. M. Molecular themes in oncogenesis. *Cell* **64:**235–248, 1991.
108. Coleman, W. B. and Tsongalis, G. J. Multiple mechanisms account for genomic instability and molecular mutation in neoplastic transformation. *Clin. Chem.* **41:**644–657, 1995.
109. Coleman, W. B. and Tsongalis, G. J. The role of genomic instability in human carcinogenesis. *Anticancer Res.* **19:**4645–4664, 1999.
110. Coleman, W. B. and Tsongalis, G. J. The role of genomic instability in the development of human cancer, in *The Molecular Basis of Human Cancer*, Coleman, W. B. and Tsongalis, G. J., eds., Humana, Totowa, NJ, pp. 115–142, 2002.

3 Extraction of Nucleic Acids

PAUL N. BOGNER AND ANTHONY A. KILLEEN

1. INTRODUCTION

The utility of DNA- or RNA-based testing depends, in large part, on the quality and nature of the diagnostic sample. A variety of methods is available to extract nucleic acids for analysis. The choice of technique should consider both the sample source and the nature of the eventual assay. Both DNA and RNA can provide valuable clinical information. Genotype analysis and infectious disease testing represent two of the primary clinical uses of DNA, but it can also be used for less obvious applications like the assessment of bone marrow transplant engraftment.

The ability to purify and test RNA provides additional and important clinical data. Purified mRNA, for example, can reveal gene expression patterns. In the last few years, two distinct types of B-cell lymphoma have been separated on the basis of their respective gene expression patterns *(1)*. Indeed, molecular classification of malignancy based on gene expression profiles is a promising and rapidly growing field. On a simpler level, the expression pattern of a single gene might be important. The *BCR/abl* fusion product seen in most cases of chronic myelogenous leukemia (CML) can be identified by reverse transcriptase–polymerase chain reaction (RT-PCR) performed on a purified RNA sample. RNA testing is also important in the detection and quantification (viral load) of retroviral infection. The utility of RNA is, unfortunately, counterbalanced by the inherently labile nature of RNA and seemingly ubiquitous presence of RNase. Nucleic acid degradation is always a concern, and RNA is especially vulnerable.

Each tissue source and extraction method presents its own potential quality assurance issues. Similarly, the needs of the eventual assay could vary. Whereas some techniques require high-molecular-weight nucleic acid (Southern blots, pulsed-field gel electrophoresis), others (including PCR-based protocols) often work well with smaller fragments. In addition to nucleic acid quality, purity and concentration are important factors to consider. Many testing techniques are sensitive to contaminating protein, lipopolysaccharide, or tissue preservative. More obvious is the need for adequate concentrations of nucleic acid. There are numerous methods for detecting and quantifying nucleic acid, with varying degrees of sensitivity and background noise. Testing based on DNA fluorescence, for example, might require higher nucleic acid concentrations than that based on radioimaging. These three parameters—quality, purity, and concentration—can be optimized with careful selection of sample source and technique.

2. SAMPLE SOURCE

There are, essentially, two types of tissue available for nucleic acid analysis: fresh and preserved. The ideal source of nucleic acid is, naturally, fresh tissue. If extraction is not possible immediately, it is critical to rapidly limit the damaging action of tissue endonucleases. Prompt flash-freezing of solid tissue with liquid nitrogen preserves nucleic acids and can facilitate subsequent tissue and cell disruption. Timely freezing is especially important to extract RNA successfully. For RNA purification, fresh tissue can also be placed directly in commercially available reagents that preserve cellular RNA for up to a week at room temperature (Ambion, Austin, TX). RNase and DNase are rapidly denatured in the presence of chaotropic agents like guanidium isothyocyanate (GITC). A minimum concentration of GITC of 5 mol/L is necessary for effective RNA preservation, and GITC-preserved tissue can also be stored at room temperature for almost a week without significant RNA loss *(2)*.

2.1. FRESH TISSUE The utility of fresh tissue is enhanced by the powerful sensitivity of PCR testing. Exfoliated cells obtained by swabbing or rinsing mucous membanes provide enough nucleic acid to allow PCR-based testing of genomic or foreign (infectious) DNA *(3,4)*. Buccal cells, for example, can be obtained noninvasively by swab and air-dried onto a glass slide. Although the cells are unfixed, the DNA is preserved well enough over short periods for transport, extraction, and PCR analysis *(3,5)*. As an alternative, buccal cells are collected with a saline mouthwash and pelleted for immediate analysis.

Cervical cells obtained by swab or brush can also be used for PCR or other DNA testing. This is especially useful in the detection of cervical human papilloma virus (HPV), subtypes of which are associated with increased risk of cervical neoplasia. A commercial system for the detection of HPV DNA has been developed that employs specific RNA probes and chemiluminescent antibody (Digene Corporation, Gaithersburg, MD).

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

Table 1
The Effect of Tissue Fixatives on the Purification of Nucleic Acid

Fixative	*Active contents*	*Tissue effect*	*Nucleic acid purification*
Neutral-buffered formalin	Formaldehyde	Nucleic acid base hydroxymethylation Crosslinking of DNA and protein	Reduced high-molecular-weight nucleic acid with increased fixation time Suitable for most testing
B-5 fixative	Mercuric chloride Usually mixed with formaldehyde	Mercury–protein complexes reduce DNA extraction yields	Low molecular weight or no extractable nucleic acid Occasionally nucleic acid sufficient for PCR testing
Bouin's fixative	Picric acid Acetic acid Formaldehyde	Acidic DNA Depurination Formaldehyde effects	Low molecular weight or no extractable nucleic acid
Zenker's fixative	Potassium dichromate Mercuric chloride	Metal–protein complexes	Low molecular weight or no extractable nucleic acid
Alcohol	Ethanol or methanol		Good or excellent nucleic acid yields, including high molecular weight Fixation time has no effect

Cervical swabs collected with this kit are useful for up to 2 wk at room temperature, and longer with refrigeration. Fixed cytologic preparations, such as Papanicolaou-stained cervical smears, can also provide useful nucleic acid for PCR testing after many years of storage *(5)*.

Fresh DNA is also available from the hair root. Again, this tissue source combines the advantages of fresh DNA with easy transport and procurement. The robust nature of the sample and PCR assay allow hair root DNA to be used in testing after proteinase K treatment, but without formal nucleic acid extraction *(3)*. This source of DNA is especially useful in forensic testing, when little other tissue might be available.

2.2. FORMALIN-FIXED, PARAFFIN-EMBEDDED TISSUE Although certainly preferable, fresh tissue is not always available for diagnostic molecular studies. For many clinical and research laboratories, the logistical limits of specimen collection make fixed solid tissue or blood more likely sources of nucleic acid.

The archival banks of formalin-fixed/paraffin-embedded tissue accumulated by most pathology departments provide a potentially vast source of tissue for both diagnostic and research analysis. Unfortunately, nucleic acid derived from fixed tissue can be less than ideal. There are many tissue fixatives currently employed by health care and research facilities. The most common is neutral-buffered formalin. Exposure of nucleic acid to formalin results in the formylation of free nucleotide amino groups, methylene bridging of bases, and crosslinking of nucleic acid with available protein. The net result is increased nucleic acid fragmentation *(6–9)*. With increased fixation time, the amount of available high-molecular-weight nucleic acid is markedly reduced. Tissue fixation in formalin for longer than 24 h will likely reduce the yield of high molecular-weight nucleic acid.

Even more problematic are fixatives containing mercuric chloride. Such fixatives include B-5 and Zenker's fixatives (both used in hematopathology). Many groups have documented limited success extracting nucleic acid from tissues fixed with these agents *(6,8–10)*. There is evidence that mercury complexes with available protein and speculation that these large complexes inhibit extraction techniques *(8–11)*. This phenomenon can be compounded with Zenker's fixative, which also contains the heavy metal chromium. Acid-containing fixatives (Bouin's fixative and Zenker's pH 2.0 fixative) could cause nucleic acid depurination. By contrast, alcohol-based fixatives allow the purification of high-quality nucleic acid. Alcohol fixation, however, is not routinely used in most applications and might not be appropriate for other clinical uses of tissue. A summary of fixative effect on nucleic extraction is presented in Table 1.

Because it is less important to have high-molecular-weight nucleic acid for most PCR applications, formalin-fixed tissue can be an excellent source of diagnostic material for these assays. Additionally, as with fresh hair or buccal samples, some PCR assays can be performed by direct amplification from formalin-fixed, paraffin-embedded tissue without prior DNA extraction *(12)*.

The problems presented by overfixation of tissue are complicated by the need for tissue without autolysis, necrosis, or inadequate fixation. With proper care and buffered formalin, most tissue specimens can be completely fixed in 24 h, and this amount of time allows for good yields of high-molecular-weight DNA *(10)*. RNA can also be extracted successfully from formalin-fixed tissue, although formalin might modify the bases enough to inhibit subsequent RT-PCR. Heating the purified sample prior to PCR might remove the offending monomethylol groups *(11)*.

2.3. BLOOD Many times, the best available DNA source is blood. Blood samples, however, present a unique problem in that the specimen is mixed with an agent to inhibit coagulation. Heparin, ethylenediaminetetraacetic acid (EDTA), and acid citrate dextrose (ACD) are all used to prevent in vitro blood clot formation. Generally, both EDTA and ACD specimen tubes provide good yields of nucleic acid appropriate for PCR and other assays. Greater than 70% of the original high-molecular-weight DNA (>25 kb) can be recovered from blood stored for 3 d in either of these preservatives, even when stored at room temperature *(13)*. Yields are even better when samples are refrigerated.

Heparin, on the other hand, is a problem when mixed with samples intended for nucleic acid extraction. Heparin adsorbs to nucleic acid and is not completely removed by standard extraction techniques. Residual heparin in a DNA or RNA sample can inhibit restriction digests, PCR, and other enzyme-based molecular biology assays. The inhibition of PCR depends, to some extent, on the relative concentrations of template and heparin. Heparin concentrations as low as 0.05 U per reaction volume might prevent amplification *(2,14)*. The sensitivity of various commercial polymerases does appear to vary, however, with some functioning normally at higher heparin levels *(14)*.

Ideally, blood samples will be obtained in either ACD or EDTA. Nonetheless, occasionally, a heparinized sample might be the only source of nucleic acid available. Attempts to remove heparin with repeated ethanol precipitation, boiling and filtering, pH modification with gel filtration, or titration with protamine sulfate do not appear to eliminate subsequent heparin assay inhibition *(15,16)*. For PCR requiring only minimal sensitivity, sample dilution might overcome this inhibition. Obviously, if amplification of a low-copy-number template is desired (e.g., infectious disease testing), sample dilution might compromise assay sensitivity. Serial washing of the buffy coat with saline prior to DNA extraction might also prove useful if white blood cells are the source of template DNA *(15)*.

In the event that a heparinized sample must be used and dilution or washes are inadequate or inappropriate, a few options remain. Heparinase treatment of the extracted DNA might allow subsequent use of the sample for high-sensitivity PCR or other testing. Heparinase is costly, however, and the heparinase preparation might be contaminated with small amounts of RNase. The presence of RNase precludes the use of heparinase in RNA purification protocols *(2)*. Alternately, heparin-free RNA can be precipitated out with lithium chloride. The addition of lithium chloride (final concentration 1.8 *M*) to a nucleic acid solution precipitates RNA, leaving inhibiting lipopolysaccharide or heparin in solution *(16)*. This technique is inexpensive and effective.

2.4. FORENSIC SAMPLES Special consideration must be given to the forensic tissue specimen. Often, these samples are neither fresh nor preserved. In general, the quantity and quality of nucleic acid decreases with specimen age. Bloodstains might provide better DNA than bone samples, especially when the specimens are old and poorly preserved *(17)*. Success has also been reported with tooth pulp, various soft tissues, and hair roots.

Complicating forensic nucleic acid degradation is environmental contamination, the unavoidable repercussion of specimen collection from an uncontrolled environment. Bloodstains, for example, can be seen on an essentially infinite variety of surfaces. Certain surface types present specific problems. Fabric dyes, especially indigo dye used in denim, could contaminate nucleic acid extractions and inhibit PCR. Using capillary action, dye can be removed by drawing saline through the fabric. Nucleic acid is transferred by this solution to a nylon membrane while dye remains in the fabric *(18)*. Other surfaces, like varnished wood, could also reduce the quality and quantity of forensic DNA *(19)*. Even on an ideal surface, the stain or sample might have been washed prior to discovery or exposed to forensic reagents like 3-aminophthalhydrazide (known as Luminol; it fluoresces in the presence of heme and is used to detect bloodstains during field investigation). Although Luminol does not appear to affect subsequent PCR analysis of extracted DNA, surface cleaning can destroy DNA evidence *(19)*. A more detailed discussion of forensic specimen collection is beyond the scope of this chapter, but it should be apparent that this is a challenging and interesting endeavor.

3. EXTRACTION

The first steps of any extraction process are tissue isolation, disruption, and cell lysis. Again, the specific protocol required depends on the sample. When a large portion of fresh or fixed solid tissue is available, it is important to select appropriate areas for subsequent harvest of nucleic acid. For collection of genomic DNA or RNA, tissue cannot be autolyzed or necrotic. Focal necrosis is common in many solid tumors, and sampling of these areas provides little or no intact nucleic acid. The histological complexity of a solid tissue sample should also be considered. Many tissues, lesions, and tumors are composed of multiple cell types and morphological areas. It is possible, and can be critical, to isolate nucleic acid from a single one of these areas. Careful dissection, and techniques like selective ultraviolet radiation fractionation *(20)* allow even small cell groups to be isolated and processed.

3.1. PARAFFIN AND BLOOD Paraffin-embedded tissues require deparaffinisation prior to nucleic acid extraction. A variety of methods exist that employ heat or solvents like xylene to remove paraffin *(21)*. Heat-based protocols are simple and require only a microwave or thermal cycler. More precise temperature control might be possible with a thermal cycler, but either system should prove successful. Some direct comparisons suggest that yields from solvent-based techniques are lower than those using heat *(22)*.

Virtually all nucleic acid obtained from blood samples, barring hematologic pathology, is leukocyte derived. Isolating white blood cells by centrifugation, therefore, could optimize purification yield and reduce reagent requirements. A simple method for this employs Ficoll lymphocyte separation medium *(21)*. Other protocols allow successful nucleic acid preparation directly from whole blood *(23)*. These techniques could prove quicker and involve fewer steps.

3.2. LYSIS/MEMBRANE DISRUPTION To purify nucleic acid from tissue samples, it is first necessary to disrupt cellular and nuclear membranes. This is efficiently accomplished with a detergent, often sodium dodecyl sulfate (SDS). The large amount of protein present in cell or nuclear lysates can make DNA or RNA purification difficult, so most methods employ proteolytic agents during this step. Proteinase K is frequently used for this purpose *(6,14,21,24)*.

3.3. ORGANIC EXTRACTION Traditional nucleic acid purification from lysate is accomplished by phenol–chloroform extraction. Variations of this basic protocol rely on the separation of protein into the organic phase and nucleic acid into an aqueous phase. It is important that the phenol pH lie within a range of 7.8–8.0 to prevent nucleic acid from remaining in the organic phase. Even at this pH, RNA with a long poly(A) tail or tract might partition with phenol. The addition of isoamyl

alcohol to the mixture prevents this, and reduces RNase activity (final ratio of phenol : chloroform : isoamyl alcohol of 25 : 24 : 1).

A key step in the organic extraction is emulsification of organic and aqueous phases. When low-molecular-weight DNA is desired, this emulsion can be achieved by vortexing. Higher-weight nucleic acid (>10 kb), however, is vulnerable to shearing forces and might tolerate only gentle shaking or rotation. The use of large-bore pipets will also reduce shearing during transfer of material, and limiting the number of transfers will also facilitate high-molecular-weight nucleic acid recovery. Extremely high-weight DNA required for pulsed-field gel electrophoresis might require cell lysis and DNA purification within an agarose plug. Digestion and removal of cellular proteins is accomplished over the course of days, leaving large and intact DNA within the agarose. This method does not employ phenol or chloroform as protein solvents.

The requirements for most molecular assays are met by conventional organic extraction, followed by ethanol precipitation. Adequate purity and yields could require serial phenol–chloroform extractions of the aqueous phase. The presence of visible protein at the interface of organic and aqueous materials warrants another round of extraction. Additionally, yield can be optimized by vigorously mixing Tris-EDTA (TE) buffer with the discarded organic phase ("back extraction"). Extra nucleic acid can be taken from the TE after subsequent centrifugation.

It is critical that the final aqueous nucleic acid solution be free of phenol and protein contamination. Care when removing the aqueous phase and repeated cycles of chloroform extraction prior to ethanol precipitation will prevent phenol contamination. In the presence of ethanol and monovalent cations, DNA or RNA precipitates out of solution at temperatures near 0°C. A variety of salts can be used as a cation source. Perhaps the most common is sodium acetate, and this is suitable for most organic extraction protocols. Other salts have unique advantages and disadvantages. Ammonium acetate reduces dNTP coprecipitation but can inhibit subsequent assays requiring nucleic acid phosphorylation. Sodium chloride is useful when samples are contaminated with SDS *(24)*. As discussed earlier, lithium chloride facilitates RNA precipitation, a technique useful in removing heparin contamination.

3.4. INORGANIC EXTRACTION As an alternative to organic purification, inorganic techniques reduce exposure to hazardous reagents while producing purified nucleic acid of comparable quality. Many commercial kits employ inorganic purification methods, including salt precipitation, adsorption to silica surfaces, and anion-exchange chromatography protocols. Many of these principles are also easily applied without commercial kits. Removal of contaminating protein by precipitation and centrifugation prior to ethanol DNA precipitation, for example, gives good yields with excellent purity *(25)*.

Especially popular are purification systems based on the binding of nucleic acid to silica or glass particles in the presence of chaotropic agents. The chaotropic agent GITC is useful for inhibiting troublesome nuclease, but it also promotes nucleic acid binding to silica or glass media *(26,27)*. Nucleic acid elution after washing can be accomplished with a low-salt aqueous buffer. Commercial systems based on this technique produce high-quality, high-purity nucleic acid preparations with improved safety and speed. There are many variations on this theme, depending on the manufacturer, and many laboratories choose to avoid the trouble of organic purification by investing in these standardized and modestly priced kits.

3.5. RNA EXTRACTION In general, variations of these techniques can be applied to both DNA and RNA purification. RNA isolation, however, demands extra care. Most forms of RNA are labile, and RNase is a frustratingly frequent contaminant of laboratory reagents and equipment. A few simple techniques and principles will help prevent degradation problems. First, reagents and equipment used for RNA preparation should be dedicated to that purpose. This becomes especially important if the laboratory also does DNA purification. RNase is frequently used in DNA work, and contamination of reagents and reusable equipment will affect subsequent RNA preparations.

Perhaps the greatest risk of RNase contamination comes from the laboratory worker's skin. It is critical that clean gloves by worn at all times and changed frequently if contact with potentially contaminated equipment is necessary.

Fresh out of the package, most sterile pipet tips and other disposable materials can be considered RNase free. Other tools might require treatment to destroy contaminating RNase. Glassware, reagent spatulas, and other equipment can be pretreated by incubation at 37°C in a solution of 0.1% diethylpyrocarbonate (DEPC). DEPC is a strong RNase inhibitor. In addition to inhibiting RNase, however, DEPC can also carboxymethylate nucleic acid purine residues. After incubation, therefore, the materials must be autoclaved to remove any remaining DEPC before use in RNA preparation. As an alternative, glassware can be baked at 150°C for 4 h or plastic materials can be soaked in 0.5 *M* NaOH for 10 min. Use of NaOH requires subsequent rinsing and autoclave treatment. Any of these methods will reduce RNase activity on reusable equipment *(21)*.

Decontamination of reagent solutions can also be accomplished by adding DEPC to a concentration of 0.1%. Caution should be exercised, however, with solutions containing amines that will react with DEPC. Tris buffer is a common example. Tris from a freshly opened or dedicated container can be added after autoclaving DEPC-treated water or solutions. To make Tris-EDTA for RNA storage, for example, a solution of EDTA can be DEPC treated and autoclaved before adding RNase-free (uncontaminated) Tris. Subsequent pH adjustments must also be made with RNase-free reagents.

3.6. DNA MICROARRAY The recent development of DNA microarray technology is also worth mentioning for its particular demands on nucleic acid extraction. In brief, cDNA transcribed from a pool of cellular mRNA is hybridized to an array of thousands of distinct DNA sequences fixed to a glass slide. Because the cDNA is transcribed with fluorescent or radioactive tags ,the relative quantity of hybridized cDNA, and therefore original mRNA, can be assessed. Generally, the gene expression profile of one cell source is simultaneously compared to a reference cell source (transcribed with a second fluorescent tag) as an internal control. This is a powerful new tool, with a rapidly growing body of literature describing its techniques and applications. One of the primary technical limits of microarray technology is the quality and quantity of RNA

available from the test and reference cells *(28)*. Contamination of nucleic acid preparations with protein, lipid, or carbohydrate can interfere with reverse transcriptase or mediate nonspecific array hybridization. A variety of mRNA purification procedures are available, and a review of public domain protocols suggests that most laboratories employ commercial kits to purify total RNA and then isolate the mRNA fraction. Total RNA can be obtained efficiently with both silica-based techniques (RNeasy; Qiagen, Valencia, CA) or precipitation procedures (TRIzol; Life Technologies, Rockville, MD). Subsequent mRNA purification is best accomplished via binding to oligo dT fixed on a solid medium or column. A number of representative protocols are available on the World Wide Web (www.microarray.org, www.nhgri.nih.gov/DIR/Microarray, and cmgm.stanford.edu/pbrown).

As genetic testing is increasingly used for clinical work, other technologies for rapid, large-scale genetic analysis will become important. The Invader® System (Third Wave Technologies, Madison, WI), for example, is an automated system that can be used to test for clinically important genetic sequences such as the Prothrombin G20210A and Factor V Leiden mutations. This system employs a combination of proprietary enzymes and specific oligonucleotide probes to generate a fluorescent signal that is amplified in a linear manner from the target DNA. No PCR amplification is required, and as little as 100 ng of target genomic DNA (or total RNA) is sufficient for testing. This method is, however, sensitive to phenol, chloroform, and high-salt concentrations, so a careful inorganic purification method is generally recommended. As more testing and detection technologies become available, the demands on purified DNA or RNA could change and will likely become less stringent.

4. ASSESSMENT OF QUALITY AND QUANTITY

In many cases, purified nucleic acid is used directly in an assay, without evaluation of its purity, concentration, or size. The results of the assay itself often demonstrate the success or failure of the preparation. In cases where assay setup or interpretation requires prior knowledge of the sample's purity and concentration, a number of methods are available.

Perhaps the simplest and fastest approach is spectrophotometry. Although nucleic acid bases show maximal absorption at an approximate wavelength of 260 nm, contaminating proteins absorb well at 280 nm. Protein absorption is primarily the result of the aromatic amino acids phenylalanine, tyrosine, and tryptophan. Testing sample optical density (OD) at 260 nm and 280 nm, therefore, allows assessment of both nucleic acid concentration and purity. An OD_{260} of 1.0 corresponds to approx 50 μg/mL of double-stranded DNA (40 μg/mL for single-stranded DNA or RNA). The $OD_{260/280}$ ratio provides an estimate of nucleic acid purity, with a pure preparation having a ratio between 1.8 and 2.0. Contamination with organic solvents or protein will obviously lower this value and also prevent accurate nucleic acid quantification from the OD_{260} reading.

If the purified sample is too contaminated or low in concentration for spectrophotometric assessment, fluorescent dyes can be used to quantify the nucleic acid present. A variety of dyes that bind nucleic acid are available, including acridine orange, daminoibenzoic acid (DABA), propidium iodide, and ethidium bromide. These dyes might detect nucleic acid in quantities as low as 1–5 ng *(24)*, although, typically, the detection threshold is closer to 5–10 ng *(29)*. Ethidium bromide, most widely used of these compounds, demonstrates fluorescence that increases by a factor of 20 or more when bound to nucleic acid. This fluorescence under ultraviolet (UV) light is directly proportional to the amount of nucleic acid present. The direct relationship between fluorescence and nucleic acid content allows the use of ethidium bromide in quantitative assays. It is worth noting that the degree of fluorescence is dependent on the ratio of ethidium bromide to nucleic acid, with maximal output at a DNA : ethidium bromide ratio of 0.5 to 3.0 (w/w) *(30)*.

Nucleic acid samples and standards (of known concentration) can be spotted with ethidium bromide on a plastic surface, and fluorescence can be captured by digital or conventional camera. Some contaminants might increase or decrease fluorescence, but this can be partly remedied by spotting on a 1% agarose slab. Small, interfering molecules will diffuse away from nucleic acid over the course of a short incubation at room temperature *(31)*.

Further information can be obtained by electrophoretic analysis of the purification product. As discussed earlier, the size of nucleic acid fragments can be important to subsequent work. Small agarose gels, "minigels," can be used easily and quickly to determine both the size and quantity of nucleic acids. For size determination, a molecular-weight ladder provides a reference standard. Likewise, for quantifying nucleic acid, a series of samples of known concentration can be run in parallel with the unknown sample, and a standard curve prepared. Again, a fluorescent dye like ethidium bromide should be used to detect nucleic acid in the gel, so that nucleic acid quantity can be extrapolated from fluorescence data.

5. NUCLEIC ACID STORAGE

When nucleic acid must be stored, either for archival purposes or before assay performance, the key goal is prevention of enzymatic or physical damage to the purified product. Three chief weapons are available for this endeavor: chelating agents, chaotrophic agents, and refrigeration. As discussed earlier, special care is needed when the solution contains RNA or high-molecular-weight DNA.

In general, DNA can be stored effectively for long periods in a Tris-EDTA buffer at 4°C. The chelation of free divalent cations by EDTA or another chelating agent diminishes the damage caused by contaminating nucleases, which require these cations for function. Cold temperatures further reduce enzyme activity and improve nucleic acid stability. RNA, inherently more labile, should be stored at −80°C in a similar buffer. As an alternative, either DNA or RNA can be stored as an ethanol precipitate, with −20°C being the optimal storage temperature.

Should the nucleic acid sample remain contaminated with nuclease after purification attempts, the addition of chaotropic reagents like GITC will deactivate remaining enzyme. GITC will crystallize at temperatures below room temperature, so solutions containing this reagent must be completely warmed before use. Likewise, RNA stored in a GITC-based solution should be kept at room temperature or frozen at −80°C.

Following these basic guidelines, purified DNA or RNA can be stored for long periods of time.

REFERENCES

1. Alizadeh, A. A., et al. Distinct types of diffuse large β-cell lymphoma identified by gene expression profiling. *Nature* **403:**503–511, 2000.
2. Neumaier, M., Braun, A., and Wagener, C. Fundamentals of quality assessment of molecular amplification methods in clinical diagnostics. *Clin. Chem.* **44:**12–26, 1998.
3. Thomson, D. M., Brown, N. N., and Clauge, A. E. Routine use of hair root or buccal swab specimens for PCR analysis: advantages over using blood. *Clin. Chim. Acta* **207:**169–174, 1992.
4. Lench, N., Stanier, P., and Williamson, R. Simple non-invasive method to obtain DNA for gene analysis. *Lancet* **i:**1356–1358, 1988.
5. Jackson, D. P., et al. Extraction of DNA from exfoliative cytology specimens and its suitability for analysis by the polymerase chain reaction. *Cytopathology* **1:**87–96, 1990.
6. Mies, C., Houldsworth, J., and Chaganti, R. S. K. Extraction of DNA from paraffin blocks for Southern blot analysis. *Am. J. Surg. Pathol.* **15(2):**169–174, 1991.
7. Dubeau, L., et al. Southern blot analysis of DNA extracted from formalin-fixed pathology specimens. *Cancer Res.* **46:**2964–2969, 1986.
8. Crisan, D. and Mattson, J. C. Retrospective DNA analysis using fixed tissue specimens. *DNA Cell Biol.* **12(5):**455–464, 1993.
9. Herbert, D. J., et al. Effects of several commonly used fixatives on DNA and total nuclear protein analysis by flow cytometry. *Am. J. Clin. Pathol.* **91(5):**535–541, 1989.
10. Koshiba, M., et al. The effect of formalin fixation on DNA and the extraction of high-molecular-weight DNA from fixed and embedded tissues. *Pathol. Res. Pract.* **189:**66–72, 1993.
11. Masuda, N., et al. Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples. *Nucleic Acids Res.* **27(22):**4436–4443, 1999.
12. Cawkwell, L. and Quirke, P. Direct multiplex amplification of DNA from a formalin fixed, paraffin wax embedded tissue section. *Mol. Pathol.* **53(1):**51–52, 2000.
13. Gustafson, S., et al. Parameters affecting the yield of DNA from human blood. *Anal. Biochem.* **165:**294–299, 1987.
14. Yokota, M., et al. Effects of Heparin on polymerase chain reaction for blood white cells. *J. Clin. Lab. Anal.* **13:**133–140, 1999.
15. Beutler, E., Gelbart, T., and Kuhl, W. Interference of Heparin with the polymerase chain reaction. *BioTechniques* **9(2):**166, 1990.
16. Jung, R., et al. Reversal of RT-PCR inhibition observed in heparinized clinical specimens. *BioTechniques* **23(1):**24–28, 1997.
17. Cattaneo, C., et al. Comparison of three DNA extraction methods on bone and blood stains up to 43 years old and amplification of three different gene sequences. *J. Forensic Sci.* **42(6):**1126–1135, 1997.
18. Larkin, A. and Harbison, S. A. An improved method for STR analysis of bloodstained denim. *Int. J. Legal Med.* **112:**388–390, 1999.
19. Gross, A. M., et al. The effect of luminol on presumptive tests and DNA analysis using the polymerase chain reaction. *J. Forensic Sci.* **44(4):**837–840, 1999.
20. Shibata, D. Extraction of DNA from paraffin-embedded tissue for analysis by polymerase chain reaction: new tricks from an old friend. *Hum. Pathol.* **25:**561–563, 1994.
21. Killeen, A. A., ed. *Molecular Pathology Protocols,* Humana, Totowa, NJ.
22. Coombs, N. J., Gough, A. C., and Primrose, J. N. Optimisation of DNA and RNA extraction from archival formalin-fixed tissue. *Nucleic Acids Res.* **27(16):**12.
23. Gustincich, S., et al. A fast method for high-quality genomic DNA extraction from whole human blood. *BioTechniques* **11(3):**298–301, 1991.
24. Sambrook, J., Fritsch, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual,* 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
25. Miller, S. A., Dykes, D. D., and Polesky, H. F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16(3):**1215, 1988.
26. Boom, R., et al. Rapid and simple method for purification of nucleic acids. *J. Clin. Microsc.* **28(3):**495–503, 1990.
27. Carter, M. J. and Milton, I. D. An inexpensive and simple method for DNA purifications on silica particles. *Nucleic Acids Res.* **21(4):**1044, 1993.
28. Duggan, D. J., et al. Expression profiling using cDNA microarrays. *Nat. Genet.* **21(1 Suppl.):**10–14, 1999.
29. Killeen, A. A. Quantification of nucleic acids. *Clin. Lab. Med.* **17(1):**1–19, 1997.
30. Dutton, M. D. and Varhol, R. J. Technical considerations for the use of ethidium bromide in the quantitative analysis of nucleic acids. *Anal. Biochem.* **230:**353–355, 1995.
31. Christen, A. A. and Montalbano, B. Estimation of quantity of polymerase chain reaction products by ethidium bromide-agarose plate assay. *BioTechniques* **9(3):**310, 1990.

4 Nucleic Acid Blotting Techniques

Theory and Practice

TERRY AMISS AND SHARON COLLINS PRESNELL

1. INTRODUCTION

This chapter deals with basic concepts and techniques in nucleic acid blotting. Many of the techniques involved with Southern blotting and Northern blotting are similar. Negatively charged, purified nucleic acids from prokaryotic or eukaryotic cells are separated according to size by electrophoresis through an agarose gel matrix. The RNA or denatured DNA is subsequently transferred and immobilized onto a membrane composed of nitrocellulose or nylon. The nucleic acids on the membrane are then hybridized to a specific labeled "probe," which consists of homologous single-stranded nucleic acids that carry molecules, allowing detection and visualization of the hybridized probe. Hybridization between the immobilized nucleic acids and labeled probe allows detection of specific DNA or RNA sequences within a complex mixture of DNA or RNA. The specific method of detection and visualization is dependent on the nature of the labeled probe; radioactive probes enable autoradiographic detection, and probes labeled with enzymes facilitate chemiluminescent or colorimetric detection. Nucleic acid blotting yields valuable information pertaining to gene integrity and copy number (Southern blot) and provides a means of analyzing gene expression and mRNA size (Northern blot). These methods can be used to characterize tissues and cultured cells in the laboratory and often provide valuable information for clinical evaluation of patient samples.

2. NUCLEIC ACID BLOTTING IN THE 21ST CENTURY

With the discovery of polymerase chain reaction (PCR) and the myriad of RNA and DNA analysis technologies that have emerged in the past decade, most scientists no longer rely solely on traditional nucleic acid analysis tools such as Northern and Southern blotting techniques. With the proper controls, the presence of a specific gene can be determined by PCR, with far less DNA required as starting material. Mutations in specific genes can also be detected through the careful design and utilization of specific primers and/or sequence analysis of amplified DNA. PCR-based methods are still challenged by the complexity of human genomic DNA, and technical limitations of PCR have hindered its use in the analysis of long stretches of DNA, such as those produced by digestion with restriction enzymes. Limitations such as these are responsible for the persistence of Southern blotting as a technical approach to DNA analysis. Ed Southern himself explains that these difficult challenges were the problem that led to the development of the first blotting method, over 25 years ago *(1)*.

The use of RT-PCR for gene expression analysis has superseded Northern blot analysis because of the development of methods that allow semiquantitative assessment of gene expression and the ability to conduct meaningful experiments with only a small amount of RNA. Although the quantity of RNA required and the time it takes to complete a Northern blot analysis make it an undesirable method for many routine studies of gene expression, there are unique advantages that Northern blotting offers. RT-PCR relies on the conversion of each copy of RNA into a corresponding cDNA, and then amplification of that cDNA into enough copies that they can be visualized; thus, there are two opportunities for mRNAs to be selected for or against: reverse transcription and amplification. Reverse transcription often selects against very small or very large mRNAs, and the success of amplification of nucleic acid relies heavily on the quality and effectiveness of the primers utilized. Furthermore, true quantitation of signal using an RT-PCR approach is more difficult than quantitation of a Northern blot signal for the above-mentioned reasons. The Northern blot has the unique advantage of detecting the true size of any mRNA and can be used to isolate novel transcripts from heterogeneous mRNA pools *(2)*. Quantitation of the signal from the Northern blot is also more straightforward, because the strength of signal is related directly to the number of copies present in the original sample and is not an extrapolated value influenced by the efficiency of the steps in the process. Thus, there are still specific situations in which traditional blotting techniques are the best choice when analyzing nucleic acids.

3. THE SOUTHERN BLOT

The DNA blot was developed by Southern in 1975 *(3)*, and it remains the method of choice among many researchers for

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

reliable, quantitative detection of specific DNA sequences. The Southern blot can detect the presence of homologous genes across species. For example, if a biologically relevant gene has been located in the rat, it is possible to construct a labeled probe from the rat gene and use it to search for a homologous gene in the human genome by Southern blot analysis. The DNA blot can also be used to assess the relative copy number of a specific gene. This particular application of the Southern blot is useful in detecting gene amplification, which is a common response to environmental pressure and often accounts for drug resistance in mammalian cells *(4,5)*. Southern analysis can identify mutations, deletions, or rearrangements that alter the integrity of a specific gene, which can be useful in the prognosis of certain types of cancer and in the prenatal diagnosis of genetic diseases. In addition, the DNA blot is a valuable tool for molecular cloning, providing a mechanism for localization of specific sequences to defined fragments within larger bacteriophage and cosmid genomic DNA clones.

4. METHODOLOGY OF SOUTHERN BLOT ANALYSIS

4.1. PREPARATION OF DNA FOR SOUTHERN BLOTTING Most basic techniques for purification of DNA produce material appropriate for Southern analysis. Standard Southern blot protocols recommend the use of 10 µg of DNA when analyzing single-copy genes *(6)*. However, when the amount of DNA is limiting, smaller quantities can be used without compromising the signal by altering the geometry of the sample well during electrophoresis (decreasing the width of the well will increase the intensity of the final signal) *(7)*. In situations where multiple copies of the gene are present or if the gene constitutes a high percentage of the DNA, such as when plasmid DNA containing the gene of interest is being analyzed, the quantity of DNA can be reduced dramatically to as little as 200 ng. Under optimal conditions, rare sequences such as single-copy genes can be detected when 10 µg of genomic DNA are analyzed *(6)*. DNA that is to be analyzed by Southern blot must first be fragmented into small pieces that can migrate through an agarose gel matrix. Restriction enzymes are bacterial enzymes that recognize specific DNA sequences (four to six nucleotides long) in DNA and cleave the DNA at these restriction sites. Digestion of genomic DNA with a given restriction enzyme produces a reproducible set of fragments that are easily separated by agarose gel electrophoresis. In order to determine which restriction enzyme(s) to use, it is helpful to know which restriction sites are present within and around the gene of interest. Generally, when evaluating the presence or copy number of a particular gene, one should avoid using restriction enzymes that cut the gene of interest into a large number of small fragments. Ideally, the gene of interest should be cleaved into a few fragments (one to three) that range in size from 1.0–10.0 kilobases (kb). A different approach might be desirable when evaluating the integrity of a specific gene. The average gene contains many restriction sites, and cleavage of the gene with a particular restriction enzyme produces a distinct number of fragments of a defined size. Mutations, deletions, or rearrangements occurring within a gene could result in a disruption of the normal nucleotide sequence, possibly altering the number of restriction sites within the gene or altering the size of the restriction fragments produced. Such a change in the size pattern of DNA fragments produced by enzymatic cleavage is referred to as a restriction fragment length polymorphism (RFLP). A detailed restriction map of a gene is generated by cleaving the DNA with several restriction enzymes separately and then performing a Southern analysis of the fragmented DNA with a probe for the gene of interest. Restriction maps are useful for identifying subtle differences between homologous genes. Hundreds of restriction enzymes are commercially available (Invitrogen [Carlsbad, CA], Sigma [St. Louis, MO], New England Biolabs [Beverly, MA]), and manufacturers typically provide the proper buffer necessary for digestion as well as instructions for the quantity of enzyme, temperature, and duration of reaction required for thorough digestion. One unit of a restriction enzyme is defined as the amount required to cleave 1.0 µg of DNA in 1 h. Complete cleavage of the DNA is essential for Southern blotting, especially when single-copy genes are being analyzed. A small aliquot (0.5–1.0 µg) of the digested DNA sample can be subjected to agarose gel electrophoresis and stained with ethidium bromide to determine whether enzymatic digestion of the DNA is complete (*see* Figs. 1 and 2). Thorough digestion of plasmid DNA is evidenced by total disappearance of the uncut plasmid and the appearance of specific bands. Digestion of genomic DNA is confirmed by the presence of a continual ladder of fragmented DNA along with distinct bands, usually in the lower portion of the gel. These bands are produced from enzymatic fragmentation of repetitive elements within the DNA and are characteristic of the restriction enzyme. After fragmentation, the DNA is typically concentrated by phenol/chloroform extraction, ethanol precipitation, and resuspended in a small volume of electrophoresis buffer in preparation for electrophoretic separation (*see* Table 1).

4.2. ELECTROPHORESIS OF RESTRICTION-DIGESTED DNA Nucleic acids are negatively charged at a neutral pH, which allows their migration through an electric field *(8)*. Agarose is a highly porous polysaccharide that acts as a sieve, allowing the fragments of DNA to be separated according to length. Under low-voltage conditions, the electrical resistance of all components remains constant and the linear DNA fragments move through the agarose gel at a velocity proportional to the voltage applied. The driving force for nucleic acid migration in the gel is the voltage gradient, which is dependent on the geometry of the electrophoresis chamber, geometry and composition of the gel, and the volume and ionic strength of the buffer used. Decreasing the distance between electrodes, decreasing gel thickness, or decreasing buffer volume can increase the velocity of DNA migration. A practical approach is to keep the gel geometry, buffer composition, and volume constant and determine the optimal running voltage empirically. The gel should be covered by 3–4 mm of buffer and high-voltage settings should be avoided, because they will lead to melting of the agarose and the appearance of artifacts on the final blot. The gel can be run overnight (12–16 h) at a low

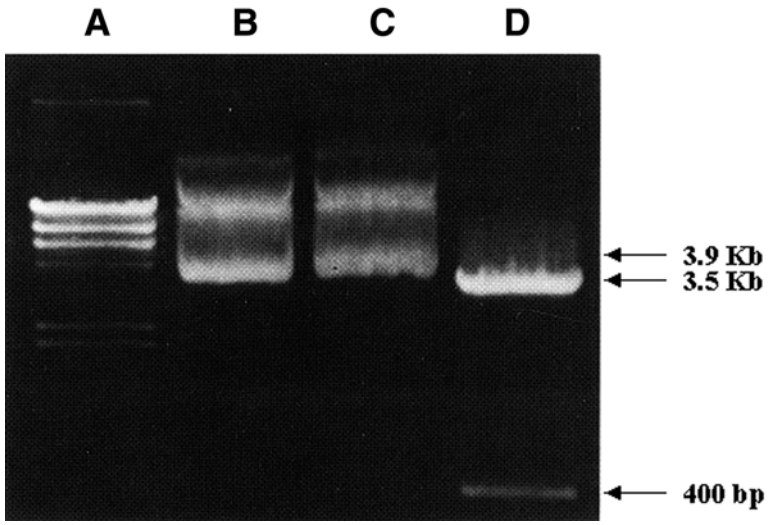


Fig. 1. Agarose gel electrophoresis of plasmid restriction digests. Five micrograms of a 3.9-kb plasmid containing a 400-bp insert of the c-*met* gene were digested with the restriction enzyme *Bst*X1 for 1 h at 37°C. Lane A contains DNA molecular-weight standards (*Hin*dIII-digested λ DNA). Lane B contains plasmid that was not incubated with a restriction enzyme. Lane C contains plasmid that was incubated at a termperature of 4°C instead of 37°C, demonstrating the importance of incubation temperature. Lane D contains fully digested plasmid DNA. Note the disappearance of the 3.9-kb DNA band corresponding to the intact plasmid, the appearance of a 400-bp DNA band representing the insert, and a 3.5-kb DNA band representing the plasmid remnant. The very high-molecular-weight band present in lanes B and C is likely to consist of aggregates of circular plasmid DNA and disappears with enzymatic digestion (lane D).

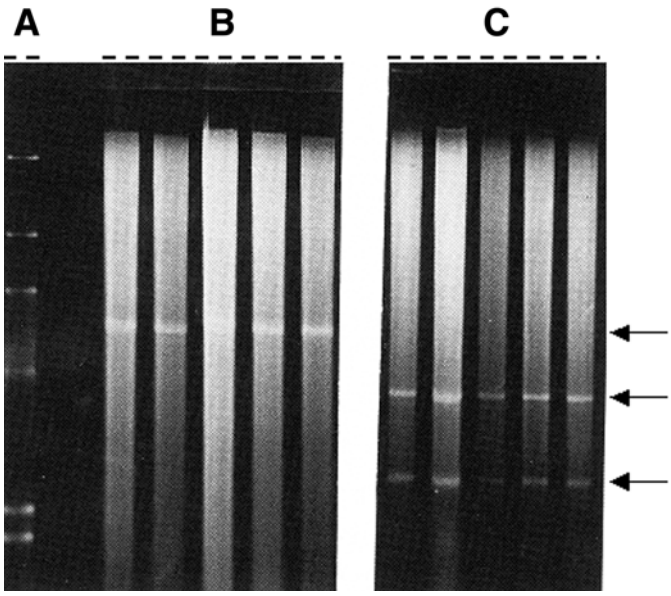


Fig. 2. Agarose gel electrophoresis of genomic DNA restriction digests. Genomic DNA samples were isolated from cultures of rat liver epithelial cells. Lane A contains DNA molecular-weight standards (*Hin*dIII-digested λ DNA). Lanes B and C show 10-µg DNA samples that were digested with either *Hin*dIII (B) or *Bam*H1 (C) for 18 h at 37°C. Agarose gel ecectrophoresis was carried out for 18 h at 22 V. The gel was stained with ethidium bromide and photographed under an ultraviolet lamp. The distinct bands (indicated with arrows) within the digested DNA represent repetitive sequences or elements in the DNA. Note that this banding pattern differs depending on the restriction enzyme employed. Thorough digestion of the DNA produces a "smear" of DNA fragments that range in size from very large (>23 kb) to very small (<0.5 kb).

voltage (20–30 V) without compromising the quality of electrophoretic separation. Because the electrophoresed DNA should be transferred to a solid support (nylon or nitrocellulose) as soon as possible, overnight electrophoresis is often a desirable option for workers with time limitations.

Agarose electrophoresis of DNA allows separation of fragments ranging from 200 to 1×10^7 basepairs (bp), although it is not possible to separate such a wide range of lengths on a single gel. A classical Southern analysis (as presented in Table 1) allows separation of fragments ranging from 200 bp to 20 kbp. Fragments smaller than 200 bp are typically analyzed by utilizing polyacrylamide gels *(9–11),* and DNA larger than 20 kbp can be analyzed by pulsed-field gel electrophoresis *(12)*.

The percentage and composition of agarose used to prepare the gel are determined based on the size of the fragment(s) of interest. Good electophoretic separation of small DNA fragments (0.2–1.0 kbp) can be accomplished using 2–4% agarose gels prepared with a 3:1 mixture of low-melting-point agarose and standard agarose (FMC BioProducts, Rockland ME). Low-melting-point agarose consists of hydroxyethylated agarose, which has better sieving properties than standard agarose and results in greater clarity of DNA bands. These high-percentage gels are useful when analyzing PCR products or cloned DNA. For Southern analysis of genomic DNA, 0.7–1.2% standard agarose gels are recommended *(7)*. The efficiency of DNA transfer to a solid support is increased with decreasing agarose concentration, but low-percentage agarose gels are delicate and difficult to manipulate. The ideal sample well size should be determined empirically. Although a weak signal can be amplified by

Table 1
Southern Blot Analysis

1. Restriction enzyme digestion
   a. Digest 10–20 µg genomic DNA with an appropriate enzyme (use 3–5 U enzyme/µg DNA).
   b. Check the efficiency of the digest by analyzing a 1-µg aliquot of DNA on an agarose gel.
   c. Precipitate the remaining digested DNA overnight with 1/10 vol of 2.5 *M* sodium acetate and 2 vol cold ethanol (100%)
   d. Resuspend precipitated DNA in 30 µL of 1X TPE and add 6 µL of of DNA sample buffer (Table 2).
2. Electrophoresis of the DNA
   a. Prepare a 0.9% agarose gel with TPE buffer (add ethidium bromide to 0.5 µg/mL).
   b. Place the gel in the electrophoresis tank and fill with TPE to 3–4 mm above gel surface.
   c. Load the samples into the sample wells; include appropriate DNA size standards.
   d. Run the gel overnight at 22–30 V (or until the bromophenol blue migrates 8 cm).
   e. Photograph the gel and carefully measure migration distances of molecular-weight standards.
3. Denaturation and neutralization of the DNA
   a. Denature the DNA by soaking the gel 2X 15 min in 0.5 *M* sodium hydroxide.
   b. Neutralize the DNA by soaking the gel 3X 10 min in 1.0 *M* Tris-HCl pH 7.5.
4. Transfer of the DNA to a nylon membrane (alkaline method)
   a. Cut a piece of nylon membrane to the exact size of the gel; prewet in dH_2O
   b. Assemble the capillary transfer apparatus as shown in Fig. 3. Take care to remove any air bubbles between the gel and the nylon membrane.
   c. Fill buffer reservoirs with alkaline transfer buffer (Table 2).
   d. Tranfer for 1–3 h (use 3 h to ensure transfer of large DNA fragments).
   e. Check the efficiency of transfer by staining the gel with ethidium bromide
   f. Let the membrane air-dry completely; *or* let the membrane air-dry briefly then fix the DNA to the filter by ultraviolet crosslinking.
5. Hybridization with labeled nucleic acid probe
   a. Prepare the probe utilizing manufacturer's instructions
   b. Prehybridize the membrane for 1 h in prehybridization solution option 1 (Table 2), at 42°C.
   c. Hybridize the membrane overnight in hybridization solution option 1 (Table 2), at 42°C.
   d. Wash the membrane 2X 15 min in 2X SSPE, 0.1% SDS, at 42°C. If additional washing is needed, wash for 30 min in 1X SSPE, 0.1% SDS, at 42°C. If necessary, subsequent washes can be performed (3X 10 min) with 0.5X SSPE, 0.1% SDS, at 42°C.
6. Visualization (radiolabeled probes): Rinse the membrane briefly in 1X SSPE, blot excess fluid from the membrane, wrap securely in plastic wrap, and expose to X-ray film for 24 h at –70°C. Develop the film and adjust exposure time as necessary.

decreasing the width of the sample well, the use of wider sample wells results in better resolution of bands.

The inclusion of a DNA size standard on analytical DNA gels containing DNA fragments of known length is recommended because such standards provide a means of extrapolating the size of a positive signal from target DNA. Formerly, a popular choice for DNA size standard in a classical Southern blot has been lambda (λ) phage DNA digested with the restriction enzyme *Hin*dIII, which provides a pattern of fragments ranging from 125 bp to 23.1 kb. Although, various DNA size standards are commercially available containing a greater number of DNA bands (Amersham [Arlington Heights, IL], Invitrogen) and can be obtained prelabeled with molecules, such as biotin, that aid in their visualization. When choosing DNA standards, it is important to be sure that the target DNA sequences are within the range of kilobase lengths represented in the DNA standards. Most protocols recommend staining the electrophoresed agarose gels with ethidium bromide to visualize the DNA standards and the digested DNA. Gels can be stained after electrophoresis by soaking in a 2-µg/mL solution of ethidium bromide. Alternatively, ethidium bromide (0.5 µg/mL) can be added to the melted agarose (after cooling to 55°C) and to the electrophoresis buffer. Staining with ethidium bromide permits photography of the gel, so the exact migration of DNA standards can be recorded along with the quality of the restriction enzyme digestion of the test DNA.

After electrophoresis, the double-stranded DNA fragments must be denatured into single strands. Denaturation of the DNA can be accomplished by soaking the gel in an alkaline solution containing sodium hydroxide (*see* Table 2). This step could be carried out on a rotary platform, which allows thorough, constant submersion of the gel. After denaturation, it is important to neutralize the gel, which is typically done by soaking the gel in a neutral (pH 7.4) solution of Tris buffer. The single strands of DNA are then ready to be transferred to a solid support, such as nitrocellulose or nylon membrane, where they can be hybridized to a complementary, labeled nucleic acid probe.

4.3. TRANSFER OF DNA TO A SOLID SUPPORT

4.3.1. Choice of Hybridization Membrane Immobilization and hybridization of nucleic acid was first carried out with nitrocellulose *(3)*. However, nitrocellulose is not ideal for nucleic acid hybridization. Because the nucleic acids are attached by hydrophobic rather than covalent interactions, they are slowly leached out of the nitrocellulose matrix during hybridization and washing at high temperatures. In addition, the fragile nitrocellulose membranes cannot survive more than one or two cycles of hybridization and washing. The shortcomings of nitrocellulose have led to the development of several alternative matrices, the most versatile of which is positively charged nylon *(13)*. Nylon membranes bind nucleic acids irreversibly and are much more durable, allowing sequential hybridizations with several different

Table 2
Solutions Used in Nucleic Acid Blotting

Solution	*Composition*	*Use*
TAE (Tris-acetate buffer)	0.04 *M* Tris-acetate, 0.1 m*M* EDTA, pH 8.0	Agarose gel electrophoresis; electrophoretic transfer of nucleic acids to a nylon membrane
TPE (Tris-phosphate buffer)	0.09 *M* Tris-phosphate, 0.2 *M* EDTA, pH 8.0	Agarose gel electrophoresis (Southern blot)
TBE (Tris-borate buffer)	0.045 *M* Tris-borate, 0.1 *M* EDTA, pH 8.0	Agarose gel electrophoresis
MOPS buffer (3-[*N*-morpholino] propanesulfonic acid)	(5X stock) 0.1 *M* MOPS, pH 7.0 40 m*M* sodium acetate 5 m*M* EDTA, pH 8.0	Electrophoresis of RNA through formaldehyde gels in Northern blotting
SSC (salt–sodium citrate buffer)	(20X stock) 3.0 *M* NaCl, 0.3 *M* sodium citrate, pH 7.0	Capillary transfer of nucleic acids to nylon or nitrocellulose membrane; washing hybridized membranes
SSPE (salt–sodium phosphate–EDTA buffer)	(20X stock) 3.6 *M* NaCl, 0.2 *M* $NaH_2PO_4 \cdot H_2O$, 20 m*M* EDTA, pH 7.7	Capillary transfer of nucleic acids to nylon or nitrocellulose membrane; washing hybridized membranes
DNA gel denaturation solution	0.5 *M* Sodium hydroxide	Denaturation of electrophoresed DNA prior to Southern blot transfer
Neutralization buffer	1.0 *M* Tris-HCl, pH 7.5	Neutralization of DNA gels after sodium hydroxide denaturation
Alkaline transfer buffer	3 *M* Sodium chloride, 8 m*M* sodium hydroxide, pH 11.4–11.45	Rapid alkaline capillary transfer of nucleic acids to nylon membranes
Denhardt's reagent	(50X stock) 1% Ficoll, 1% polyvinylpyrrolidine, 1% bovine serum albumin	A blocking agent added to hybridization solutions in Southern and Northern blots to reduce background
Prehybridization solution (option 1)	5X SSPE, 5X Denhardt's reagent, 200 µg/mL denatured salmon sperm DNA, 0.1% SDS, 50% formamide	Prehybridzation of Northern and Southern blots prior to hybridization with labeled probe
Hybridization solution (option 1)	5X SSPE, 2.5X Denhardt's reagent, 200 µg/mL denatured salmon sperm DNA, 0.1% SDS, 50% formamide	Hybridization of Northern and Southern blots to labeled probe
Prehybridization and hybridization solution (option 2)	6X SSC, 2X Denhardt's reagent, 0.1% SDS	Prehybridization and hybridization of Northern and Southern blots
Prehybridization and hybridization solution (option 3)	5X Denhardt's reagent, 0.5% SDS, 100 µg/mL denatured salmon sperm DNA	Useful for reducing high-background hybridization in Northern and Southern blots
DNA and RNA sample buffer	50% Glycerol, 1 m*M* EDTA, 0.25% bromophenol blue, 0.25% xylene cyanol FF	Electrophoresis of RNA through formaldehyde gel

probes without a loss of membrane integrity. Nylon membranes also allow highly efficient electrophoretic transfer of small amounts of nucleic acid when capillary or vacuum transfer is insufficient. The single disadvantage of nylon membranes is the propensity for higher-background hybridization, but this problem can be eliminated by increasing the amount of blocking agents during prehybridization and hybridization.

4.3.2. Methods of DNA Transfer Single-stranded (denatured) DNA can be transferred to a solid support, such as nitrocellulose of nylon membrane. If nitrocellulose membranes are used, the recommended method of transfer is capillary transfer. With the more versatile nylon membrane, several methods of transfer are available, including capillary transfer, electrotransfer, and vacuum transfer. Although nitrocellulose can be used for vacuum transfer or electrotransfer, these methods are optimal with nylon membranes. Regardless of the method of transfer, it is recommended that the gel and the membrane be equilibrated in transfer buffer prior to transfer. The composition of buffer is dependent on the method of transfer employed (*see* Table 2).

In capillary transfer (1), nucleic acid fragments are eluted from the gel and deposited onto the membrane by transfer buffer that is drawn through the gel by capillary action. Rate of transfer is dependent on the size of the fragments, with larger fragments transferring less efficiently. The disadvantage of traditional capillary transfer is the length of time required for efficient transfer of large nucleic acid fragments which transfer less efficiently (usually overnight). When large (>5 kb) fragments of DNA are to be analyzed, many protocols recommend depurinating the electrophoresed DNA prior to denaturation by soaking the gel in 0.2 *M* hydrochloric acid for 5–15 min *(14)*. Depurination, along with denaturation, leads to the breakdown of long DNA fragments into shorter pieces, which transfer more efficiently. However, this nicking of the DNA has been reported to reduce the final hybridization signal significantly and decrease the clarity of bands on the autoradiograph *(5)*. Downward alkaline capillary transfer, which can be completed

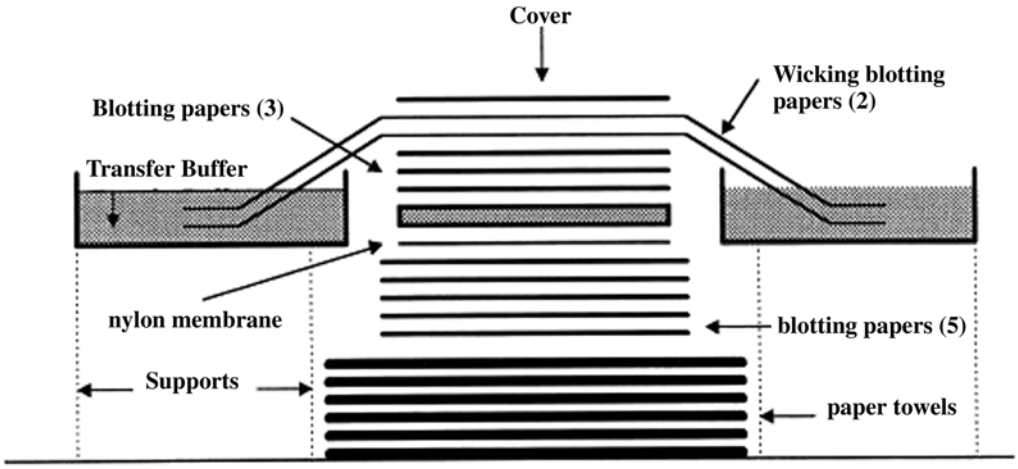


Fig. 3. Downward capillary transfer of nucleic acids to hybridization membranes.

in 1–3 h, offers a fast and efficient alternative to traditional capillary transfer *(15,16)* and does not require special equipment (*see* Fig. 3).

Electrophoretic transfer *(17)* was developed as a faster alternative to capillary transfer. Size-separated nucleic acid fragments are transferred onto a membrane (preferably nylon) by placing the gel between porous pads that are inserted between parallel electrodes in a large buffer tank. These apparatuses are available from several manufacturers (Bio-Rad [Hercules, CA], Schleicher & Schuell [Keene, NH]). Electroblotting can be quite efficient, with complete transfer of high-molecular-weight nucleic acids in 2–3 h. However, the electrophoresis apparatus must be equipped with a cooling mechanism for maintaining an acceptable buffer temperature, and large buffer volumes are required. The need for special equipment has limited the practical application of the electroblot to situations in which capillary transfer or vacuum transfer is not sufficient.

Vacuum blotting was introduced as another alternative to capillary transfer *(18,19)*, and involves the application of negative pressure to the gel. Nucleic acids are eluted by buffer that is drawn through the gel by application of negative pressure (a vacuum). Transfer of nucleic acids by vacuum blotting is more efficient than capillary transfer and has been reported to result in a twofold to threefold increase in final hybridization signal obtained *(20)*. Vacuum blotting devices are commercially available (Bio-Rad) and work well when the vacuum is applied evenly over the gel surface. However, efficiency of transfer could be reduced if the vacuum exceeds 60 cm of water *(6)*, because of compression of the gel. When carried out properly, the vacuum blot provides a means for fast (approx 4 h), efficient, highly reproducible transfer of nucleic acids.

4.3.3. Fixing DNA onto Nitrocellulose or Nylon Membranes After transfer, the DNA must be adhered stably to the membrane to ensure that it remains in place during hybridization and washing. If a nitrocellulose membrane was used, the DNA can be affixed by baking the damp membrane at 80°C for 2 h in a vacuum oven. Alternatively, nylon membranes can be exposed (DNA side up) to low-level ultraviolet (UV) irradiation at 254 nm. Irradiation of the membrane results in crosslinks between the nucleic acid residues and positively charged amine groups on the membrane surface *(21)*. Overirradiation of the membrane could cause covalent attachment of a high percentage of the nucleic acid residues, resulting in a decreased hybridization signal. Special ovens for the irradiation of membranes are commercially available, and most manufacturers recommend 1.5 J/cm^2 for damp membranes and 0.15 J/cm^2 for dry membranes. Optimally, the ideal amount of irradiation should be determined empirically. Baked or irradiated Southern blots can be stored at room temperature until they are ready to be hybridized to a labeled nucleic acid probe.

5. INTERPRETATION OF THE SOUTHERN BLOT

Once the membrane that contains the target DNA sequences has been "probed" with a specific labeled probe, the results of the Southern analysis can be interpreted. Positive signals on the membrane are created when a labeled nucleic acid probe binds to complementary target DNA sequences, producing a band or bands that can be visualized. If nonradioactive, colorimetric probing techniques were used, the bands would be visible to the naked eye. Chemiluminescent and radioactive probes must be visualized by exposing the hybridized membrane to X-ray film, producing an autoradiograph. Figure 4 shows an autoradiograph generated by hybridization with a ^{32}P-labeled radioactive probe. The autoradiographs that are produced by Southern blotting are easily analyzed by a scanning densitometer—an instrument that provides a measure of the relative density of the bands that are present on the X-ray film. The end result is a numerical value, which can be used to determine the relative number of target sequences present in one sample compared to another. For example, the relative density of a band produced by hybridization with a gene whose copy number has been amplified will be much higher than the relative density of the band corresponding to the normal gene copy number. Membranes that have been hybridized with radioactive or chemiluminescent probes can also be analyzed directly in a phosphorimager (Amersham Biosciences, Piscataway, NJ). This instrument

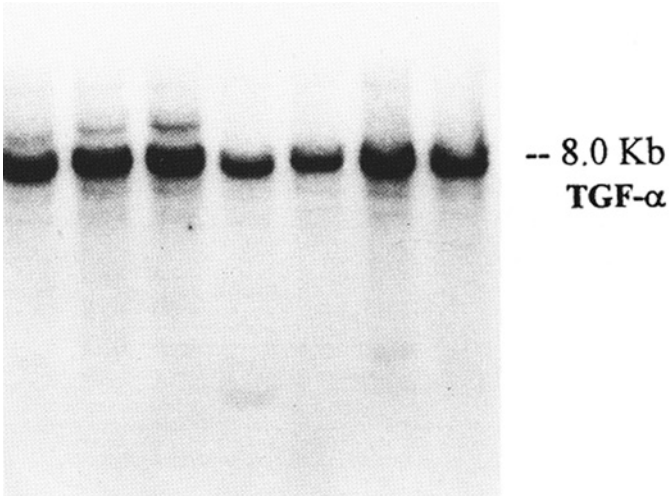


Fig. 4. Southern blot analysis of transforming growth factor (TGF-α). DNA was isolated from cultures of rat liver epithelial cells. Ten-microgram DNA smaples were digested with the restriction enzyme *Bam*H1 for 18 h at 37°C. The digested DNA was subjected to Southern blot analysis as described in Table 1. The ^{32}P-labeled probe was generated by random primer extension, utilizing a 1.8-kb fragment of the rat TGF-α gene. The membrane was hybridized, washed, and exposed to X-ray film for 48 h. When rat liver epithelial cell DNA is digested with *Bam*H1, the TGF-α gene is cut into an 8.0-kb fragment, which hybridizes to the radiolabeled probe and appears as a dark band in the 8.0-kb range on the autoradiograph.

provides a very accurate evaluation of the blot by measuring the amount of radioactive chemiluminescence emission corresponding to each band on the membrane, generating a value expressed in counts per minute. The scanning densitometer and phosphorimager enable the user to evaluate the positive signals on a blot with concrete numerical data instead of guesswork.

Proper interpretation of Southern blots necessitates the use of controls throughout the procedure. In order to compare the gene copy number between one sample and another, it is imperative that the same quantity of DNA be loaded onto the gel for all samples being compared. This can be assured by measuring the DNA concentration carefully prior to loading and by evaluating the relative concentrations of DNA on the electrophoresed gel by staining with ethidium bromide after electrophoretic separation. When analyzing patient samples that are suspected to contain abnormalities (gene deletions, rearrangements, or amplifications), one should include DNA that is known to be normal for comparison.

6. TROUBLESHOOTING SOUTHERN BLOTS

There are several problems that may occur that will result in the failure of a signal to appear on the final blot. To help eliminate possibilities, consult the following checklist:

Was enough DNA loaded onto the gel (10 μg)?
Was the DNA digested thoroughly prior to electrophoresis?
Was the electrophoresed DNA denatured and neutralized before transfer?
Was transfer complete? (Stain the gel after transfer to assess residual DNA.)
Was the DNA immobilized properly onto the membrane support?
Was the DNA intact? (Has the membrane been stripped and probed multiple times?)
Was the probe prepared properly?
Was the hybridization time sufficient (at least overnight)?
Was the blot exposed to film for a sufficient amount of time?
Was the probe labeled sufficiently?

If the answer to all of the above questions is "yes," then the failure of a positive signal to appear could be related to the strength with which the probe hybridized to the target sequences on the membrane. If the probe is from a species other than the test DNA, weak hybridization might be the result of a lack of homology between the probe and the target gene sequence. Decreasing the stringency of the posthybridization washes and/or increasing the length of exposure time while generating the autoradiograph can often strengthen a very weak signal. Alternatively, it might be necessary to try a different probe or a different probe-labeling technique.

High background on the final blot is usually indicative of insufficient blocking during the prehybridization and hybridization steps or of poor washing of the blot after hybridization. To eliminate the background, simply increase the amount of denatured salmon sperm DNA added to the prehybridization and hybridization solutions and/or increase the stringency of the posthybridization washes by decreasing salt concentration or increasing temperature. When RNA probes (riboprobes) are utilized, the temperature of hybridization washes might need to be as high as 65°C to eliminate background hybridization.

7. THE NORTHERN BLOT

The Northern blot allows identification of specific messenger RNA sequences within a mixture of RNA molecules. The final signal achieved on the blot is proportional to the number of specific sequences present, allowing for a quantitative analysis of gene expression. The RNA transcripts produced from a particular gene can vary in size because of phenomena such as the following:

1. Utilization of a secondary transcription start site by RNA polymerase;
2. Premature termination of transcription resulting from nonsense mutation;
3. Posttranscriptional modifications, such as splicing;
4. Deletions within the gene coding sequence.

These alternative transcripts can be detected by Northern blot analysis. This information is routinely used to determine if the expression of a specific gene is altered in any way. Aberrations in gene expression are frequently studied in the laboratory to evaluate the cellular response to a particular stimulus or treatment at the molecular biology level. Because abnormal expression of specific genes is often reliably associated with certain disease states, the Northern blot can also be a valuable tool for diagnosis/prognosis in the clinical setting.

8. GENERAL CONSIDERATIONS IN THE ANALYSIS OF RNA

8.1. THE IMPORTANCE OF AN RNAASE-FREE ENVIRONMENT Successful Northern blot analysis requires a preparation of RNA that is intact and relatively free from contaminants, such as protein and DNA. When working with RNA, it is also imperative to maintain an environment that is free of Rnases (ubiquitous enzymes that degrade RNA). This can be accomplished by treating deionized water and solutions that contact the RNA with 0.1% diethylpyrocarbonate (DEPC) for 24 h at 3°C prior to autoclaving *(6,13)*. Solutions containing Tris (tris[hydroxymethyl]aminomethane) cannot be treated with DEPC. Therefore, it is recommended to reserve a bottle of Tris that is used only for RNA work and is handled appropriately. Sterile, RNAase-free tubes and pipet tips should be used at all times, and glassware should be baked or rinsed with chloroform to eliminate RNAases *(6)*. Wash solutions are commercially available that destroy Rnases and could be useful in cleaning glassware and work stations (RNaaseAWAY, Invitrogen). The abundance of Rnases in skin necessitates the use of gloves by personnel handling RNA samples, solutions, and labware that comes in contact with samples.

8.2. TOTAL CELLULAR RNA VS MESSENGER RNA There are many variables to take into consideration when determining whether to analyze total cellular RNA or purified messenger RNA. The specificity and efficiency of the specific probe and the level of expression of the target gene contribute to the strength of the final signal observed on a Northern blot. In a preparation of total cellular RNA, only about 2.5% is actually mRNA; the majority is composed of ribosomal RNA (rRNA). The quantity of RNA that can be subjected to Northern analysis is limited to approx 30 µg by technical limitations related to the capacity of hybridization membranes for RNA binding. Therefore, detection of rare messenger RNAs (mRNAs) frequently requires utilization of purified mRNA. There are many commercially available kits for purification of mRNA, all of which take advantage of the polyadenylation signal on the 3′ end of mRNAs. The basic principle involves selection of the mRNA by binding the sequential adenosine residues to a synthesized stretch of deoxythymidine residues that have been affixed to a solid support (such as cellulose or magnetic beads). The unbound ribosomal RNA and other contaminates are washed away, and the mRNA is then eluted from the support matrix. Even rare mRNAs can be detected when as little as 1 µg of purified RNA is subjected to Northern analysis.

9. METHODOLOGY OF NORTHERN BLOT ANALYSIS

9.1. PREPARATION OF THE RNA FOR ELECTROPHORESIS The method of choice for RNA isolation in many labs is centrifugation through a cesium chloride cushion, followed by ethanol precipitation *(22)*. This method efficiently generates RNA of exceptional quality for Northern blotting. However, alternative methods are available that address limitations of some investigators with respect to instrumentation. The acidic phenol extraction method described by Chomczynski *(23)* yields total cellular RNA of acceptable quality for Northern blot analysis, can be performed in less time than a cesium chloride gradient, and does not require an ultracentrifuge. Commercial kits are available for the isolation of RNA, most of which are based on variations of acidic phenol extraction.

Regardless of the method utilized to generate total cellular RNA, it is advisable to check the quantity and quality of the samples by taking spectrophotometer readings at 260 and 280 nm and running a small aliquot (1 µg) through a 1.0% agarose integrity gel containing 0.5 µg/mL ethidium bromide. The ratio of the absorbance at 260/280 is 1.8–2.0 in a clean RNA sample. Ratios <1.8 indicate contamination with protein or phenol. Visualization of the integrity in a gel with a UV light source is the best measure of RNA quality (*see* Fig. 5). DNA contaminants are revealed as very high-molecular-weight bands that sometimes fail to migrate into the gel. Degraded RNA is identified as a smear in the very low-molecular-weight range. When total cellular RNA is analyzed, 1S and 28S rRNA bands should be clearly visible, and a faint smear representing the heterogeneous mRNA population should be present as well. The eukaryotic 28S and 18S rRNAs are 5.0 and 1.87 kb in size, respectively *(7)*, and can serve as a convenient internal RNA size standard.

9.2. ELECTROPHORETIC SEPARATION OF RNA A protocol for Northern blot analysis is provided in Table 3. When analyzing total cellular RNA, the quantity of RNA loaded onto the gel can range from 5 to 30 µg. The ideal concentration should be determined empirically and depends on the quantity and quality of target mRNA as well as specificity of the probe. Analysis of purified mRNA can be performed with as little as 1.0 µg, although a greater quantity (5–10 µg) is typically used to permit multiple uses of a single membrane. Just prior to electrophoresis, the single-stranded native RNA molecules must be denatured to abolish secondary structure (*see* Table 2). This is accomplished by heating the samples to 65°C in the presence of formaldehyde and formamide. The denatured state is maintained during electrophoresis by the addition of formaldehyde to the agarose gel *(24)*. Denaturation of RNA with glyoxal was introduced in 1977 as an alternative to formaldehyde denaturation *(25)*; although glyoxal denaturation works well, additional steps are required to remove the glyoxal products after blotting. The standard RNA gel (1.2% agarose, 1.1% formaldehyde) allows separation of RNA from 0.5 to 6.0 kbp. It is advisable to run RNA molecular weight standards on every gel to aid in determining the molecular weight of bands present. Because RNA and DNA do not migrate at the same rate, DNA standards are not acceptable for RNA gels. RNA standards are commercially available in several size ranges. The addition of ethidium bromide to RNA gels is controversial. Although visualization of the RNA provides information on integrity and quantity, experimental evidence suggests that the subsequent transfer of RNA to nitrocellulose or nylon is impeded when ethidium bromide is present, resulting in a 12–18% decrease in final hybridization signal *(13,26)*. For this reason, many researchers remove and stain only the lane containing the RNA standards. However, the valuable

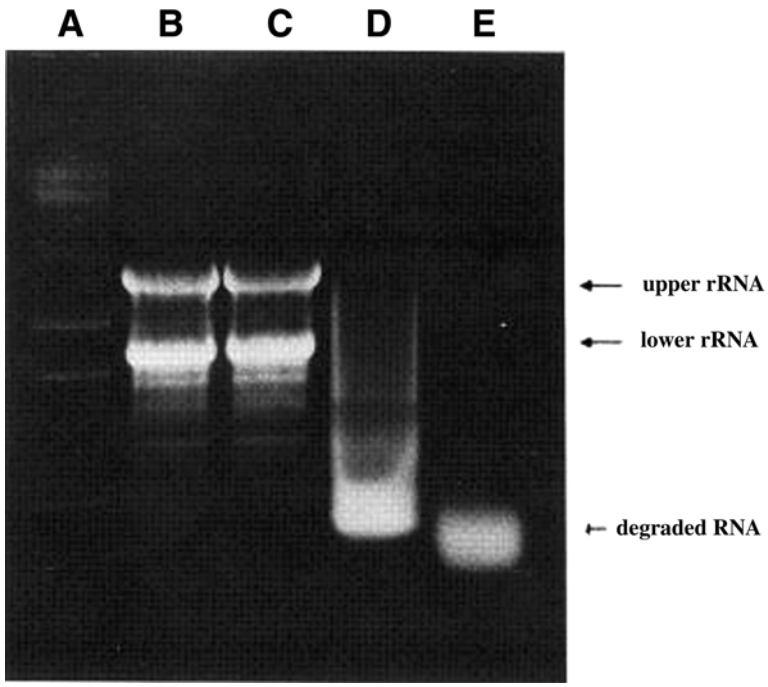


Fig. 5. RNA integrity gel. RNA was isolated from cultured rat liver epithelial cells by the acidic phenol extraction method. Two micrograms of RNA per sample were run on an integrity gel for 2 h at 90 V. Lane A contains an RNA molecular size standard. Lanes B and C contain intact RNA (rRNA bands are clearly visible), whereas the samples in lanes D and E are degraded. The RNA in lane D was prepared in a small volume of tap water (instead of RNAase-treated water) and the RNA in lane E came in contact with human skin during sample preparation.

information obtained from visualization of the RNA often outweighs the inconvenience of a reduced signal. Alcidine orange *(25)* and Stains All (a cationic carbocyanine dye) *(25,27)* are alternative dye choices for staining and visualizing the RNA.

The recommended buffer for electrophoresis of RNA contains 3-(*N*-morpholino)propanesulfonic acid (MOPS) *(6)*. A 5X or 10X stock of MOPS buffer (*see* Table 2) should be prepared in DEPC-treated water, brought to a pH of 7.0, filter-sterilized, and stored at room temperature in a dark bottle. Autoclaving or exposure to light can cause yellowing of the buffer. Although pale yellow buffer could still be used, buffer exhibiting a darker shade of yellow should not be used. Constant circulation of buffer during electrophoresis in the same direction as electrical current prevents accumulation of buffer components and formaldehyde in the positive buffer chamber. The general rule for running RNA gels is 3–4 V/ cm length. The bromophenol blue (in the RNA sample buffer) should migrate at least 8 cm before electrophoresis is terminated. Because there is no reliable storage method for electrophoresed RNA gels, it is optimal to transfer the nucleic acids to a solid support immediately after electrophoresis. For this reason, many workers choose to run gels overnight at a low voltage (20–30 V).

If ethidium bromide is not included in the gel, the lane containing RNA standards must be excised and stained. The distance from the loading well to each band should be measured to generate a standard curve. A plot of the $\log_{10}$ of the RNA fragment sizes against distance migrated provides a curve by which sizes of RNA species detected by hybridization can be calculated. Prior to transfer of the RNA to a nitrocellulose or nylon membrane, the formaldehyde should be removed from the gel. This is typically accomplished by soaking the gel in several changes of DEPC-treated water or transfer buffer.

9.3. TRANSFER OF RNA TO A SOLID SUPPORT For optimal results, the electrophoresed RNA should be transferred to a nylon membrane. The methods of transfer (capillary, vacuum, and electrophoretic) are the same as for DNA transfer and are discussed in previous sections. Because the RNA molecules are already single-stranded, there is no need to denature the gel prior to transfer. However, both the gel and the nylon membrane should be equilibrated in the appropriate transfer buffer (*see* Table 2) prior to transfer. Traditional capillary transfer works well for most Northern blots, although complete transfer takes 12–18 h (*see* Fig. 3). When time is limited, the downward alkaline transfer or vacuum transfer method can be used. Electrophoretic transfer is ideal when the target mRNA is present in a very low quantity or when large RNAs need to be transferred. When fixing a Northern blot, experimental evidence indicates that fixation by UV irradiation is superior to baking at 80°C *(21)*. As with DNA, overirradiation can cause extensive crosslinking of the RNA and diminish the final signal, so optimal UV exposure should be determined empirically. Irradiated, air-dried membranes can be stored at room temperature until they are ready to be hybridized to a nucleic acid probe.

9.4. INTERPRETATION OF THE NORTHERN BLOT When analyzing any experimental sample by Northern blot, it is recommended to include a control on the gel consisting of

Table 3
Northern Blot Analysis

1. Preparation of RNA samples
   a. Combine an appropriate quantity of purified total RNA or mRNA with 2.0 µL of 5X MOPS buffer, 3.5 µL of 37% formaldehyde, pH > 4.0, 10.0 µL formamide, and RNAase-free water to give a total volume of 20 µL.
   b. Add 4.0 µL of RNA loading dye (Table 2) and heat samples to 65°C for 10 min. Open the caps for 5–10 min prior to loading the samples to allow traces of ethanol to evaporate.
2. Preparation of the formaldehyde gel
   a. Melt agarose (final concentration should be 1.2%) in RNAase-free water by boiling in a microwave. Cool to 60°C, add 5X MOPS gel buffer (1/5 vol) and 37% formaldehyde (1/5.6 final volume). If desired, ethidium bromide can be added to the gel (0.5 µg/mL).
   b. Pour gel immediately into clean, level gel tray and insert the appropriate sample comb(s).
   c. When the gel has solidified (approx 20 min), place it in the electrophoresis tank, remove the sample comb, and fill the tank with 1X MOPS buffer until the liquid covers the gel by 3–4 mm. Prerun the gel for 5 min at 5 V/cm length.
3. Running the gel
   a. Load the RNA samples carefully into the sample wells. Reserve one well for loading appropriate RNA size standard (1–3 µg are usually sufficient). If the samples "float" out of the well as they are loaded, they might still contain traces of ethanol. Reheating the samples briefly (uncapped) will eliminate this problem.
   b. The gel may be run overnight at 20–30 V, or at 3–4 V/cm length. Electrophoresis should continue until the bromophenol blue tracking dye has migrated a minimum of 8.0 cm.
   c. Once the samples have entered the gel (approx 20 min), begin circulating the buffer with a peristaltic pump in the same direction as the electrical flow.
   d. When electrophoresis is complete, record the migration distances of the RNA ladder (if ethidium bromide was not added to the gel, this lane should be cut away and stained in a 2-µg/mL solution of ethidium bromide). If ethidium bromide was added to the gel, the entire gel should be visualized with a UV light source and photographed. Remove the lane containing the RNA standards prior to transfer.
4. Capillary transfer of the RNA to a nylon membrane
   a. Cut a piece of nylon membrane to the size of the gel and prewet it in RNAase-free water. Equilibrate the membrane in alkaline transfer buffer for 15 min.
   b. Remove the formaldehyde from the gel by soaking 3X 10 min in RNAase-free water.
   c. Assemble the transfer setup as shown in Fig. 3 and allow the RNA to transfer for 1–3 h.
   d. Air-dry the nylon membrane completely; *or* air-dry the membrane briefly than crosslink with UV irradiation.
5. Hybridization with nucleic acid probe
   a. Prehybridize the membrane for at least 1 h at 42°C in 10–20 mL of prehybridization solution.
   b. Prepare the labeled probe as per manufacturer's instructions.
   c. Replace prehybridization solution with hybridization solution containing the labeled probe and incubate 12–14 h at 42°C.
6. Washing the hybridized membrane
   a. Wash the membrane 2X 15 min in 2X SSPE, 0.1% SDS at 42°C, and then 2X 15 min in 1X SSPE, 0.1% SDS at 42°C.
   b. If a radioactive probe was used, check the membrane with a Geiger counter. If additional washing is needed, wash 2X 15 min in 0.5X SSPE, 0.1% SDS at 42°C.
7. Visualization of hybridized probe (radiolabeled probes): Wick excess buffer from the washed membrane and wrap the filter in plastic wrap; expose the membrane to film in a lighttight cassette at –70°C. Develop the autoradiograph after 24 h, adjusting the exposure time as needed.

RNA isolated from a control cell or tissue. Interpretation of a Northern blot is most accurate when actual numerical values are assigned to the bands present on the final blot, which represent a positive signal. Colorimetric detection methods are the least sensitive and should be avoided if comparisons are to be made between two test samples in which differences in expression might be subtle. Both chemiluminescent and radioactive probes can be exposed to X-ray film to produce an autoradiograph. Figure 6 shows an autoradiograph generated from a Northern blot that was probed with a ^{32}P-labeled cDNA probe. The autoradiograph can be analyzed by a scanning densitometer, which measures the density of each band and assigns a numerical value to the band. This analysis allows the detection of very small variations in expression. However, for interpretation to be accurate, a control must be included for the amount of mRNA that is present in the samples being compared. Some researchers rely on ethidium bromide staining for the gel prior to transfer and hybridization. However, visualization of the mRNA in formaldehyde-containing gels is difficult and not quantitative. A more reliable method is to perform a second hybridization reaction with the membrane, utilizing a probe (labeled in the same way as the probe used to detect the gene of interest) for a "housekeeping gene," such as actin (a structural protein) or cyclophilin (cyclosporine-binding protein). Numerical values are then generated for the positive signal from the housekeeping gene via scanning densitometry. The expression of the gene of interest is then expressed as a function of the expression of the housekeeping gene, thereby "normalizing" each signal relative to the actual quantity of mRNA present in that particular lane of the gel. If radiolabeled or chemiluminescent probes are used, the blots can be analyzed directly in a phosphorimager, which detects positive signals on the blot and expresses the value as actual counts per minute of radioactivity. This instrument is particularly useful in detecting very weak positive signals and can often shorten the time required to detect positive bands compared to generating traditional autoradiographs.

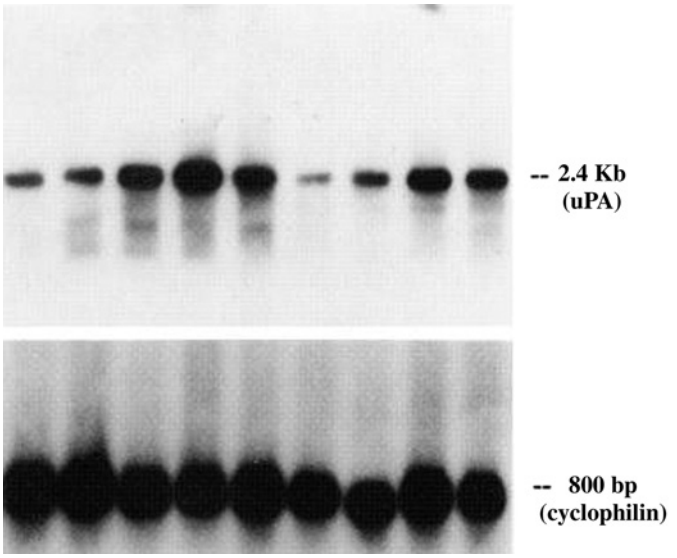


Fig. 6. Northern blot analysis of the urokinas plasminogen activator (uPA) mRNA. RNA was isolated from nine different rat liver epithelial cell lines by acidic phenol extraction and was subsequently purified by binding to oligo-dT. Two-microgram samples of mRNA from each cell line were subjected to Northern blot analysis as described in Table 3. The radiolabeled probes were generated by random primer extension utilizing a 1.5-kb fragment of the uPA gene and a 1.2-kb fragment of the cyclophilin gene. The blot was first probed with the probe for uPA and an autoradiograph was generated. Then, the membrane was stripped in a solution containing 50% formamide and 2X SSPE and reprobed for the cyclophilin mRNA. Although the expression of the "housekeeping" gene (cyclophilin) remains relatively constant among the samples, the expression of the 2.4-kb uPA mRNA varies greatly among this group of cell lines.

10. TROUBLESHOOTING NORTHERN BLOTS

Some genes are simply not expressed at any level in certain cell types. However, if a known positive control was included on the gel and no positive signal was produced, there could be a technical problem. Consult the following checklist to eliminate some possibilities:

Was the RNA of good quality (integrity gel, ethidium bromide staining)?
Was electrophoresis of the RNA sufficient (migration of RNA bp ladder)?
Was the RNA transferred to the nylon membrane (stain the gel after transfer)?
Was the probe labeled properly?
Was hybridization time sufficient (overnight)?
Was the blot washed too stringently?

If no technical problems can be identified, the best solution is to reprobe the blot with a probe for a known housekeeping gene. If no signal is present, the RNA might have been degraded or present in too small a quantity to be detected.

A high-background signal might be reduced by increasing the amount of blocking agent (denatured salmonsperm DNA) in the prehybridization and hybridization reactions. However, background hybridization is frequently the result of insufficient washing of the blot after hybridization. Washing should be carried out in generous volumes (at least threefold volume of hybridization solution), and increasing the stringency of the wash (increasing temperature and/or decreasing salt concentration) will reduce background signal. A washing routine that works well for one probe might be insufficient for another, so optimal washing conditions must be determined separately for each probe used.

Sometimes, a positive band on a Northern blot creates a great deal of excitement until it is found to be the wrong size. Such occurrences lend a lot of support to the use of molecular-weight standards, such as the RNA bp ladders, to prevent misinterpretation of data. If pure mRNA has been used and an unexpected band appears, it is possible that an alternative transcript has been identified. When mRNA is not pure (as when total cellular RNA is analyzed), nonspecific binding of probe to the rRNAs can be observed. For this reason, one should take great care in interpreting Northern blots when rRNA is present. Any bands that appear at 5.0 or 1.87 kb could be attributed to binding of the probe to rRNA rather than hybridization with true target sequences. In situations in which these problems arise, it is advisable to repeat the blot utilizing highly purified mRNA.

11. PREPARATION OF LABELED NUCLEIC ACID PROBES

Once nucleic acids have been affixed to a membrane, specific sequences can be detected by hybridization with a labeled, denatured, single-stranded probe that binds to homologous RNA or DNA. These probes might be composed of either RNA or DNA, and labeling methods might be radioactive or nonradioactive.

11.1. NICK TRANSLATION The method of nick translation relies on *Escherichia coli* DNA polymerase I—a polymerizing enzyme that also possesses a 5′→3′ exonuclease activity that degrades double-stranded DNA and RNA : DNA hybrids *(26)*. First, discontinuities ("nicks") are generated in the phosphodiester backbone of the double-stranded DNA by brief treatment with pancreatic DNase I, producing free 3′-hydroxyl termini along the strand of DNA. DNA polymerase I then extends the 3′-OH termini in the presence of the four dNTPs, utilizing its exonuclease activity to hydrolyze nucleotides in the 5′→3′ direction (*see* Fig. 7). The use of radioactive nucleoside triphosphates in the reaction with DNA polymerase I produces uniformly labeled DNA. Disadvantages of nick translation include the strict requirements in the protocol to time and temperature limitations and the large amount of template DNA (0.5 μg) required per reaction. It is also important to note that small DNA fragments (<200 bp) are not suitable for nick translation. Nick translation kits are commercially available and provide a reliable source for the buffers and enzymes to carry out the reaction. Most protocols recommend separating radiolabeled DNA from unincorporated dNTPs by centrifugation through a small column of Sephadex G-50 *(6,28)*.

11.2. RANDOM PRIMER EXTENSION An alternative method for generating labeled DNA involves the utilization of oligonucleotide primers of random sequence *(29)*. The double-stranded DNA is denatured and random nanomers or hexamers

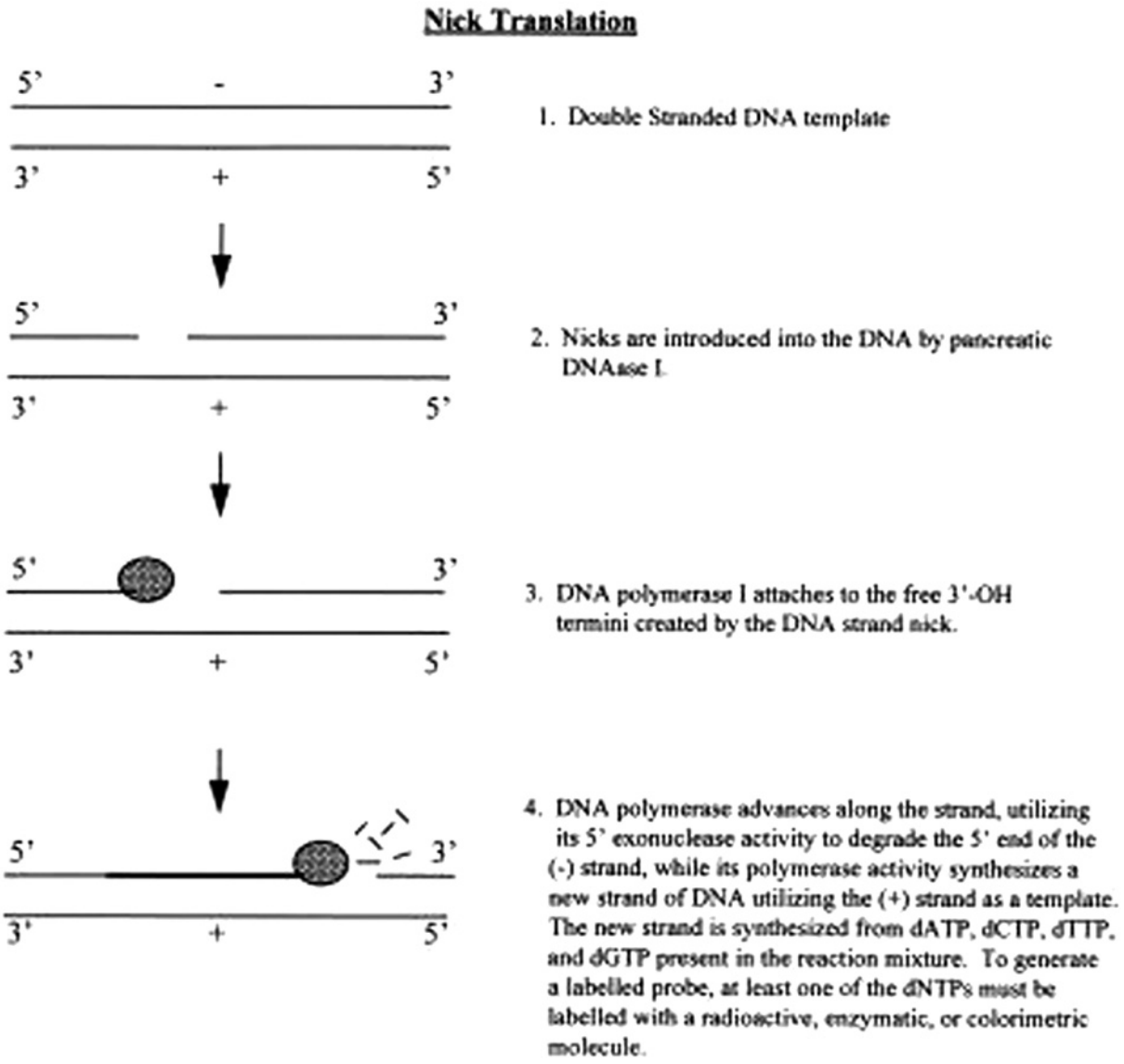


Fig. 7. Preparation of DNA probes by nick translation.

are annealed to the template DNA strands (*see* Fig 8). The primers are extended by the large fragment of DNA polymerase (Klenow fragment) or T7 DNA polymerase in the presence of radiolabeled dNTPs. Random-primed probes can be labeled to a higher specific activity than nick-translated probes, although they are typically shorter (approx 500 nucleotides). The reaction can also be carried out at room temperature with small quantities of template DNA (25–50 ng) and is not significantly affected by longer incubation times (overnight). This method could also be used to label small fragments (200 bp). Random primer extension kits are available commercially and provide the user with enzyme, cold dNTPs, and reaction buffers. Because the majority of radioactive dNTP is incorporated into DNA, purification of the probe is not usually necessary. However, if purification is needed, centrifugation through Sephadex G-50 is sufficient.

11.3. GENERATION OF STRAND-SPECIFIC PROBES

The generation of radiolabeled probes from double-stranded DNA works well when the target sequences are present in sufficient quantity and the hybridization between the labeled DNA probe and target sequences is strong. When hybridization between probe and target is weak, hybrids form between the complementary DNA sequences of the probe, resulting in segregation of the probe and decreased detection of target sequences. Single-stranded probes are composed of only one of two strands of a nucleic acid sequence, thus allowing detection of target sequences without unwanted reannealing of probe. These probes are particularly useful when analyzing target sequences that are only partially homologous to the probe (such as the detection of homologous genes in multiple species).

Radiolabeled cDNA probes are generated by primer extension of single-stranded DNA derived from a recombinant bacteriophage M13 and can yield probes of extremely high specific activity [1×10^9 counts per million (CPM)/μg] *(29)*. Primers are commonly chosen that anneal to the single-stranded viral DNA in a region upstream from the site of insertion. Extension of the annealed primers is typically accomplished with the Klenow fragment of DNA polymerase I

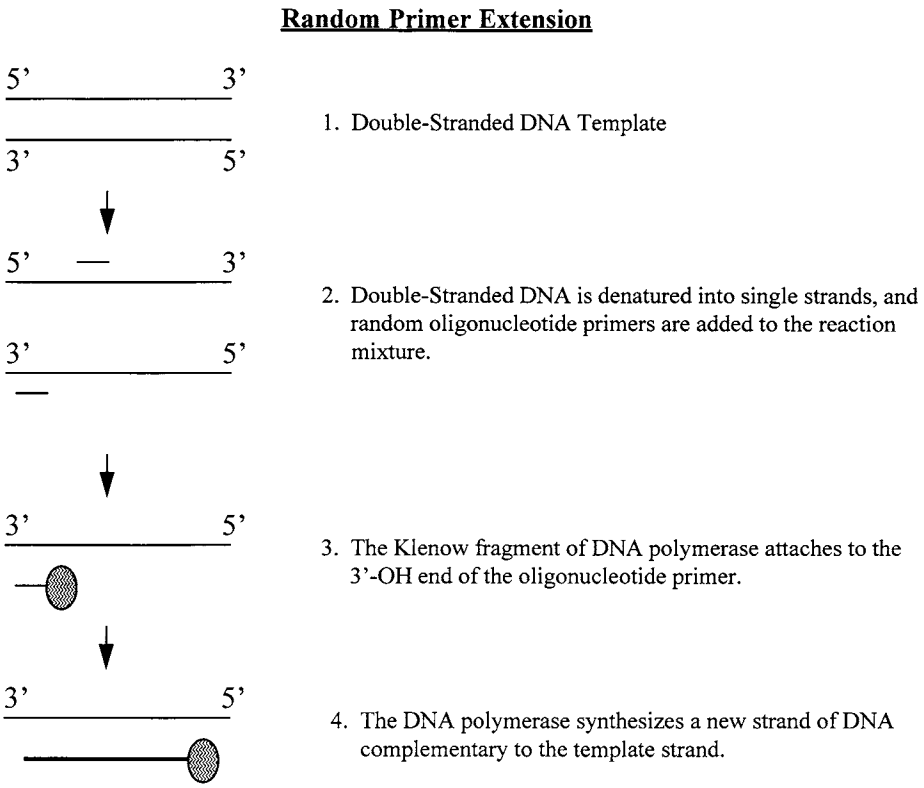


Fig. 8. Preparation of DNA probes by random primer extension.

in the presence of three nonradioactive dNTPs and one α-^{32}P-labeled dNTP. The major drawback of this method is the required separation of labeled probe from the template and smaller DNA fragments. Labeled probes can be separated by polyacrylamide gel electrophoresis or by alkaline chromatography through Sepharose CL-4B (Pharmacia Biotech) *(28)*.

The ability to generate single-stranded RNA probes ("riboprobes") has been made possible by the development of plasmid vectors, which contain multiple cloning sites downstream from strong bacteriophage promoters (SP6, T7, and T3). These promoters are recognized by bacteriophage-specific, DNA-dependent RNA polymerases, which fail to recognize bacterial, plasmid, or eukaryotic promoters that are present within the construct *(30)*. As a result, labeled RNA can be synthesized directly from these linearized plasmids at an extremely high efficiency and specific activity when radioactive NTPS are included in the reaction (*see* Fig 9). The greater stability of RNA hybrids makes RNA probes superior to double-stranded and single-stranded DNA probes in Southern and Northern analysis. In addition, the production of radiolabeled RNA is more efficient than the generation of single-stranded cDNA probes, because unwanted template DNA can be eliminated by simply treating the samples with RNase-free DNase I. When these RNA probes are to be applied to Northern blots, it is important to generate the probe in the "antisense" direction, so that the probe is complementary and not identical to the mRNA of interest.

11.4. GENERATION OF NONRADIOACTIVE PROBES

Environmental concerns, cost, and safety are a few reasons that nonradioactive alternatives to ^{32}P-labeled probes are becoming more popular. Another advantage of nonradioactive probes is their long half-life, which permits them to be stored for extended periods of time. Several vendors provide kits for nonradioactive labeling and detection of probes, all of which employ the same basic concepts. Probes can be labeled by the traditional enzymatic methods (described earlier for radioactive probes), with the incorporation of alternatively labeled dNTPs or NTPs. Nonradioactive probes are typically labeled with haptens (such as digoxigenin, biotin, or fluorescein) *(31–33)*. Detection of the hybridization signal depends on the type of label used, but it is either colorimetric or chemiluminescent.

Biotin-labeled probes are usually detected by enzyme-conjugated streptavidin and by enzyme-conjugated antihapten antibodies that detect hapten-labeled probes. The conjugated enzymatic activity (horseradish peroxidase or alkaline phosphatase) results in the production of a colored precipitate in the presence of a specific chemical substrate *(34)*, which allows direct visualization of hybridized bands on the membrane. However, colorimetric detection has several disadvantages that must be considered. Detection of rare sequences is limited and often requires long development times (>15 h). Furthermore, the colored precipitate cannot be efficiently removed from the membrane, thus preventing multiple uses of a single blot.

Chemiluminescent detection relies on the association of an enzyme-conjugated antibody with the digoxigenin, biotin, or fluorescein moieties of labeled probes. The reporter enzyme (horseradish peroxidase or alkaline phosphatase) dephosphorylates a chemiluminescent substrate, thus generating an unstable anion that emits light as it decomposes *(35)*. Positive signals are visualized by exposing the blot to X-ray film for short

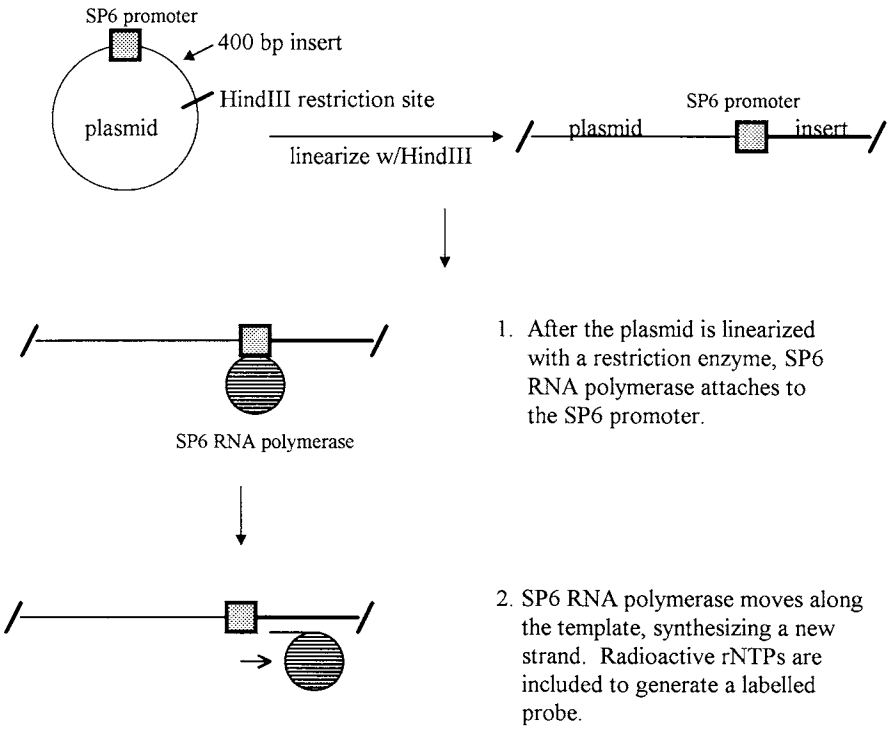


Fig. 9. Generation of riboprobes.

lengths of time. Chemiluminescence does not produce a precipitate, which allows blots to be stripped and reprobed by standard methods. Chemiluminescent detection is sufficiently sensitive for most applications and is currently the most accepted alternative to radioactive probes. The major drawback to chemiluminescent detection is the frequent observation of high-background activity. However, strict adherence to the blocking steps provided in most protocols can significantly reduce background signal.

12. HYBRIDIZATION WITH LABELED PROBES

12.1. FACTORS INFLUENCING HYBRIDIZATION The kinetics of nucleic acid hybridization rely heavily on such parameters as temperature, salt concentration, solvent concentration, and the relative strength of nucleic acid hybrids (RNA : RNA > RNA : DNA > DNA : DNA) *(6)*. Blots are typically prehybridized (at the hybridization temperature) in a solution that reduces nonspecific binding of labeled DNA during hybridization. The prehybridization solution is then replaced with hybridization solution, which contains the labeled probe. Composition of prehybridization and hybridization solutions can vary depending on the protocol and subsequent detection method. Two commonly used formulations are presented in Table 2. The temperature of hybridization largely determines the specificity of the signal obtained in Northern and Southern analysis (high temperatures increase stringency). Many protocols recommend the use of formamide-based prehybridization and hybridization solutions to lower the required hybridization temperature without compromising stringency. Hybridization can be carried out in sealed plastic bags, although commercially available hybridization bottles and ovens are recommended as a safe and reliable alternative. Some protocols recommend the use of high-molecular-mass polyethylene glycol or dextran sulfate in the hybridization solution to concentrate the probe and enhance the final signal *(6,7)*.

12.2. WASHING AND VISUALIZATION OF NUCLEIC ACID BLOTS After hybridization the, nonspecifically bound probe is removed by sequential washes with buffer (*see* Tables 1 and 3). The temperature and composition of the wash buffer have an effect on the specificity of the signal obtained. Higher temperatures and/or lower salt concentrations increase the stringency of the wash and remove the probe that is not strongly hybridized to complementary sequences. When radioactive probes are utilized, the washing process can be easily monitored with a Geiger counter. Visualization of hybridized radiolabeled probes is accomplished by exposing the membrane to X-ray film. The use of tungstate-based intensifying screens can intensify a positive signal 10-fold when applied at –70°C.

12.3. REMOVAL OF BOUND PROBES FROM NUCLEIC ACID BLOTS If nitrocellulose or nylon membranes do not dry out during the blotting process, they can be stripped and reprobed. Probes can be removed from nitrocellulose by immersing the membrane in hot elution buffer (*see* Table 2) *(6)*. Stripping of nylon membranes (DNA or RNA) can be accomplished using one of the following:

1. Immersing the filter in 1m*M* Tris-HCl, 1m*M* EDTA, and 0.1X Denhardt's reagent for 2 h at 75°C;
2. Immersing the membrane in a formamide-based stripping solution containing 2X SSPE and 50% formamide for 1 h at 65°C;
3. Pouring boiling water over the membrane and letting it stand for 5–10 min.

Probes can also be removed from DNA blots by alkaline treatment with sodium hydroxide at 42°C, followed by neutralization in Tris buffer at 42°C *(7)*. Once the probe is removed, the hybridization membrane should be air-dried briefly and sealed in a plastic bag. Sealed blots can be stored at 4°C for extended periods of time without deterioration.

13. SUMMARY

Nucleic acid blotting yields valuable information in both the research and clinical settings. More than 500 human genetic diseases are attributed to single-gene defects. For example, sicklecell anemia is caused by a simple point mutation in the gene that encodes for the β-chain of hemoglobin *(36)*. By employing short oligonucleotide probes for mutant and normal gene sequences, it is possible to distinguish between the two forms of the gene by Southern blot analysis *(37)*. Pathological gene alterations can also be detected in diseases such as Burkett's lymphoma, Fragile X syndrome, familial hypercholesterolemia, the hemophilias, and inborn errors of metabolism *(38,39)*. The Southern blot provides a means of identifying a specific altered gene, thus allowing prenatal detection of many genetic diseases, many of which can be corrected by simply administering the functional protein that should be encoded by the defective gene. The Northern blot provides information on gene expression that might be useful in establishing prognosis in diseases, such as colon cancer *(40)*. During embryonic development, large numbers of genes are switched "on" and "off" in elaborate, defined patterns. This if often true for the process of carcinogenesis as well, in which the expression patterns of specific genes become altered, contributing to the formation of a tumor. Northern blotting allows the expression of a specific gene to be examined so that one can determine if the gene of interest is underexpressed or overexpressed compared to normal. Experimental evidence supports the idea that carcinogenesis is the result of the loss of a cell or group of cells to control their growth. Such a phenomenon might be the result of overproduction of a growth-stimulatory factor, underproduction of a growth inhibitor, or altered production of cell cycle regulatory molecules.

Nucleic acid blotting permits detailed characterization of a specific gene. Southern blot identify structural abnormalities, whereas Northern blots show gene expression levels and low detection of alternatively spliced transcripts. Despite the development of more rapid PCR-based methods of nucleic acid analysis, traditional nucleic acid blotting remains a valuable tool in many molecular biology laboratories.

REFERENCES

1. Southern, E. M. Blotting at 25. *TIBS* **25:**585–588, 2000.
2. Porchet, N. and Aubert, J.-P. Northern blot analysis of large mRNAs, *in Glycoprotein Methods and Protocols: The Mucins,* Methods in Molecular Biology Vol. 125, Corfield , A., ed., Humana, Totowa, NJ.
3. Southern, E. M. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98:**503–517, 1975.
4. Ruiz, J. C., Choi, K. H., von Hoff, D. D., Roninson, I. B., and Wahl, G. M. Autonomously replicating episomes contain mdr1 genes in a multidrug-resistant human cell line. *Mol. Cell. Biol.* **9:**109–115, 1989.
5. Stark, G. R. and Wahl, G. M. Gene amplification. *Annu. Rev. Biochem.* **53:**447–491, 1984.
6. Sambrook, J. F., Fritsch, E. F., and Maniatis, T., eds., *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory, Press, Cold Spring Harbor, NY, 1992.
7. Kroczek, R. A. Southern and northern analysis. *J. Chromatogr.* **618:**133–145, 1993.
8. Bostian, K. A., Lee, R. C., and Halvorson, H. O. Preparative fractionation of nucleic acids by agarose gel electrophoresis. *Anal. Biochem.* **95:**174–182, 1979.
9. Knowland, J. S. Polyacrylamide gel electrophoresis of nucleic acids synthesised during the early development of *Xenopus laevis* Daudin. *Biochim. Biophys. Acta* **204:**416–429, 1970.
10. Preat, T. High resolution southern analysis of genomic DNA using heat denatured acrylamide gels. *Nucleic Acids Res.* **18:**1073, 1990.
11. Vesterburg, O. A. A short history of electrophoretic methods. *Electrophoresis* **14:**1243–1249, 1993(abstract).
12. Burmeister, M. and Ulanovsky, L. *Pulsed-Field Gel Electrophoresis*, Humana, Totowa, NJ, 1992.
13. Van Oss, C. J., Good, R. J., and Chaudhury, M. K. Mechanism of DNA (Southern) and protein (Western) blotting on cellulose nitrate and other membranes. *J. Chromatogr.* **391:**53–65, 1987.
14. Wahl, G. M., Stern, M. and Stark, G. R. Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulfate. *Proc. Natl. Acad. Sci. USA* **76:**3683–3687, 1979.
15. Chomczynski, P. One-hour downward alkaline capillary transfer for blotting of DNA and RNA. *Anal. Biochem.* **201:**134–139, 1992.
16. Ming, Y. Z., Di, X., Gomez-Sanchez, E. P., and Gomez-Sanchez, C. E. Improved downward capillary transfer for blotting of DNA and RNA. *Biotechniques* **16:**58–59, 1994.
17. Smith, M. R., Devine, C. S., Cohn, S. M., and Lieberman, M. W. Quantitative electrophoretic transfer of DNA from polyacrylamide or agarose gels to nitrocellulose. *Anal. Biochem.* **137:**120–124, 1984.
18. Kroczek, R. A. and Siebert, E. Optimization of northern analysis by vacuum-blotting, RNA-transfer visualization, and ultraviolet fixation. *Anal. Biochem.* **184:**90–95, 1990.
19. Stacey, J. and Isaac, P. G. Restriction enzyme digestion, gel electrophoresis, and vacuum blotting of DNA to nylon membranes. *Methods Mol. Biol.* **28:**25–36, 1994.
20. Olszewska, E. and Jones, K. Vacuum blotting enhances nucleic acid transfer. *Trends Genet.* **4:**92–94, 1988.
21. Khandjian, E. W. UV crosslinking of RNA to nylon membrane enhances hybridization signals. *Mol. Biol. Rep.* **11:**107–115, 1986.
22. Kornguth, S. E., Anderson, J. W., Scott, G., and Kubinski, H. Fractionation of subcellular elements from rat central nervous tissue in a cesium chloride gradient. Biochemical and ultrastructural studies. *Exp. Cell Res.* **45:**656–670, 1967.
23. Chomczynski, P. and Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Anal. Biochem.* **162:**156–159, 1987.
24. Lehrach, H., Diamond, D., Wozney, J. M., and Boedtker, H. RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. *Biochemistry* **16:**4743–4751, 1977.
25. McMaster, G. K. and Carmichael, G. G. Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. *Proc. Natl. Acad. Sci. USA* **74:**4835–4838, 1977.
26. Kroczek, R. A. Immediate visualization of blotted RNA in northern analysis. *Nucleic Acid. Res.* **17:**9497, 1989.
27. Wade, M. F. and O'Conner, J. L. Using a cationic carbocyanine dye to assess RNA loading in Northern gel analysis. *Biotechniques* **12:**794–796, 1992.
28. Hagel, L., Lundstrom, H., Andersson, T., and Lindblom, H. Properties, in theory and practice, of novel gel filtration media for

standard liquid chromatography. *J. Chromatogr.* **476:**329–344, 1989.
29. Kaguni, J. M. and Kaguni, L. S. Enzyme-labeled probes for nucleic acid hybridization. *Methods Biochem. Anal.* **36:**115–127, 1992.
30. Butler, E. T. and Chamberlin, M. J. Bacteriophage SP6-specific RNA polymerase. I. Isolation and characterization of the enzyme. *J. Biol. Chem.* **257:**5772–5778, 1982.
31. Yamaguchi, K., Zhang, D., and Byrn, R. A. A modified nonradioactive method for northern blot analysis. *Anal. Biochem.* **218:** 343–346, 1994.
32. Dubitsky, A., Brown, J., and Brandwein, H. Chemiluminescent detection of DNA on nylon membranes. *Biotechniques* **13:**392–400, 1992.
33. Murakami, A., Tada, J., Yamagata, K., and Takano, J. Highly sensitive detection of DNA using enzyme-linked DNA-probe. 1. Colorimetric and fluorometric detection. *Nucleic Acids Res.* **17:**5587–5595, 1989.
34. Nakagami, S., Matsunaga, H., Oka, N., and Yamane, A. Preparation of enzyme-conjugated DNA probe and application to the universal probe system. *Anal. Biochem.* **198:**75–79, 1991.
35. Schaap, A. P., Akhavan, H., and Romano, L. J. Chemiluminescent substrates for alkaline phosphatase: application to ultrasensitive enzyme-linked immunoassays and DNA probes. *Clin. Chem.* **35:**1863–1864, 1989.
36. Ingram, V. M. A specific chemical difference between the globins of normal human and sickle-cell anemia hemoglobin. *Nature* **178:**792–794, 1956(abstract).
37. Chang, J. C. and Kan, Y. W. A sensitive new prenatal test for sickle-cell anemia. *N. Engl. J. Med.* **307:**30–32, 1982.
38. Burkitt, D. and O'Connor, G. T. Malignant lymphoma in African children. A clinical syndrome. *Cancer* **14:**258–269(abstract).
39. Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds. *The Metabolic Basis of Inherited Disease, 6*th ed., McGraw-Hill, New York, 1989.
40. Fearon, E. R. and Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61:**759–767, 1990.

5 The Polymerase Chain Reaction

William B. Coleman and Gregory J. Tsongalis

1. INTRODUCTION

It is difficult to underestimate the impact of the polymerase chain reaction (PCR) and related DNA amplification techniques on modern molecular biology and applied molecular medicine. PCR represents a rapid, sensitive, and specific method for amplification of nucleic acid sequences and is the basis for numerous molecular techniques that have become the mainstay of the basic research laboratory, as well as the clinical diagnostics laboratory. The concept of PCR was first described in 1985 *(1)*, and the modern technique emerged a few years later *(2)*. Since that time, the technology has evolved into a reliable and affordable method that is performed in laboratories worldwide.

When reduced to its essence, PCR is a molecular technology that facilitates the amplification of rare copies of specific nucleic acid sequences to produce a quantity of amplified product that can be analyzed. In early descriptions of PCR *(1,3,4)*, the Klenow fragment of *Escherichia coli* DNA polymerase I was used for DNA synthesis during each amplification cycle. However, Klenow fragment is not thermal-stable. Therefore, after each denaturation step the samples were quickly cooled before the addition of enzyme to avoid heat denaturation of the polymerase enzyme and it was necessary to add a fresh aliquot of Klenow fragment enzyme after each denaturation cycle. In addition, the primer hybridization and DNA synthesis steps were carried out at 30°C to preserve the activity of the polymerase enzyme, resulting in hybridization of primers to nontarget sequences and considerable nonspecific amplification *(4)*. Even with these drawbacks, the original PCR methodology was successfully applied to gene cloning and molecular diagnostic experiments *(1,3,4)*. The major technological breakthrough in development of PCR came with the introduction of a thermostable polymerase to PCR *(2)*. *Thermus aquaticus* is a bacterium that lives in hot springs and is adapted to the variations in ambient temperature that accompany its environment. The DNA polymerase enzyme expressed by *T. aquaticus* (known as *Taq* polymerase) exhibits robust polymerase activity that is relatively unaffected by rapid fluctuations in temperature over a wide range *(5)*. Introduction of *Taq* to PCR improved the practicality of this methodology. Because *Taq* polymerase can survive extended incubation at the elevated temperatures required for DNA denaturation (93–95°C), there is no need to add a new enzyme after each cycle. In addition, by using a heat block that automatically changes temperatures (a thermocycler), the PCR cycles becomes automated. Incredibly, the basic PCR technique has not changed that much since 1988 *(2)*, although new developments in commercially available molecular reagents have made the technique easier to perform.

In this chapter, we review basic concepts related to the PCR, its reaction components, and contemporary molecular methods that are based on PCR. The information presented is not intended to represent or function as a laboratory manual for PCR applications. However, excellent laboratory manuals on PCR (and related technologies) are available for the interested reader *(6)*.

2. POLYMERASE CHAIN REACTION

In a typical PCR, successive synthetic cycles are performed in which DNA polymerase copies a target DNA sequence from a template molecule in vitro (Fig. 1). The amplification products of each cycle provide new templates for the next round of amplification. Thus, the concentration of the target DNA sequence increases exponentially over the course of PCR. The typical PCR mixture contains (1) a thermostable DNA polymerase (*Taq* polymerase), (2) target-specific forward and reverse oligodeoxynucleotide primers, (3) each of the four deoxynucleotide triphosphates (dNTPs), (4) reaction buffer, and (5) a source of template (genomic DNA, cDNA, or a cell lysates). The target sequence is defined by the specificity of the oligodeoxynucleotide primers that anneal to complementary sequences on opposite template strands flanking the region of interest. During the PCR, these primers are extended in the 5′→3′ direction by the polymerase enzyme to yield overlapping copies of the original template. Each cycle of PCR proceeds through three distinct phases: (1) denaturation, (2) primer annealing, and (3) primer extension. The denaturation step is typically accomplished by incubation of samples for 1 min at 94°C to unwind the DNA containing the gene of interest. The primer annealing step is accomplished at a temperature that is specific for the PCR primers and conditions employed, typically 1 min at 55–65°C.

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

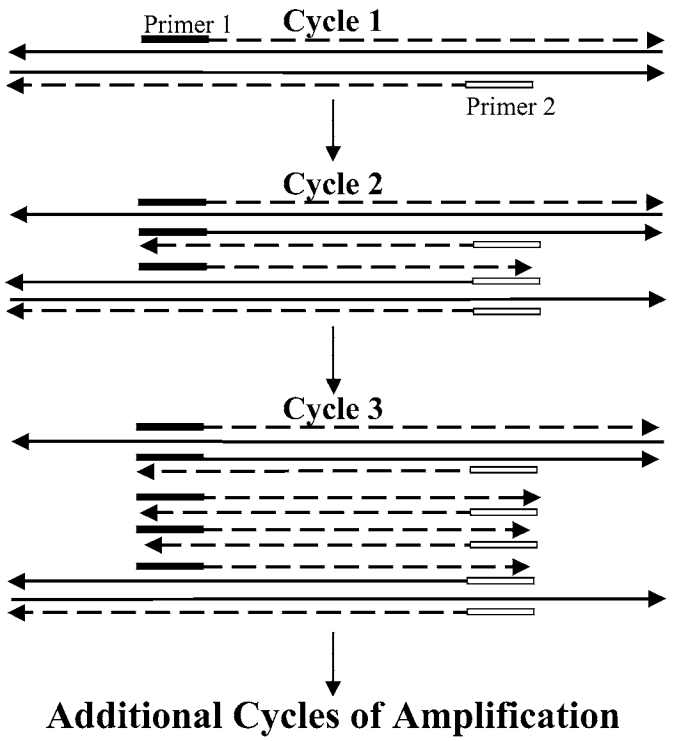


Fig. 1. Polymerase chain reaction. In each cycle of PCR, sequence-specific primers flanking a target sequence anneal to template DNA, and primer extension is accomplished using *Taq* polymerase (or other thermostable polymerase enzyme). In the first cycle of PCR, the template is composed entirely of input sample (genomic DNA or cDNA). With the second and subsequent cycles of PCR, newly synthesized PCR products become templates for additional amplification. The amplified products accumulate exponentially during a typical 25- to 35-cycle PCR.

During the annealing step, oligodeoxynucleotide primers recognize and hybridize to the target sequence contained within the single-stranded template. The primer extension step is accomplished at 72ºC (for *Taq* polymerase) for 1–2 min. During this step, the polymerase enzyme catalyzes the polymerization of dNTPs in a DNA-directed DNA synthesis reaction. The actual times used for each cycle will vary (from 15 s to 1–2 min) depending on the type of thermocycler employed and its temperature ramping speed. Amplification of target sequence is accomplished through repetition of these incubations for 25–30 (or more) cycles. The exact number of cycles necessary to produce an amplicon that can be detected (visualized or otherwise detected) will depend on the starting concentration of the target sequence. By the end of the third cycle of amplification, a new double-stranded molecule is formed in which both the 5′ end and 3′ end coincide exactly with the primers employed (Fig. 1). Because the copy number theoretically doubles after each successive cycle, an exponential increase of 2^n (n is the total number of cycles) is accomplished during the complete chain reaction. Accumulation of amplicons corresponding to the target sequence eventually reaches a plateau. The initial number of target sequences contained within the template sample, the efficiency of primer extension, and the number of PCR cycles performed determine the upper limits of amplification. The majority of double-stranded products formed during subsequent cycles of PCR are of a defined size and are seen clearly as a sharp band when analyzed by gel electrophoresis.

3. PCR COMPONENTS

3.1. NUCLEIC ACID TEMPLATES FOR PCR Polymerase chain reaction amplifies specific sequences from DNA templates (either genomic DNA or cDNA derived from RNA) that can be prepared from various sample sources (Fig. 2). This DNA could be genomic DNA isolated directly from experimental or patient material or it could be cDNA that has been synthesized from DNA or RNA templates by polymerase or reverse transcriptase enzymes. Various sources of biological material can be utilized for the preparation of PCR templates. In the research lab, PCR templates can be derived from cultured cells or tissue sections. Clinical specimens could be derived from various bodily fluids (such as blood and amniotic fluid), as well as surgical samples (such as frozen tumors) *(7)*. Forensic specimens could be derived from hair samples, blood, or semen *(8)*. In addition to fresh specimens, DNA derived from fixed tissues (paraffin-embedded specimens) can be used routinely (with rare exception) in PCR applications *(9)*. In fact, PCR analysis has been applied to prehistoric DNA derived from fossilized biological materials *(10)*.

Most PCR reactions amplify small targets in the DNA sequence (200–1000 bp). Thus, high-molecular-weight DNA is not necessary, and highly fragmented DNA (like that obtained from paraffin tissue sections) can be effectively utilized. However, certain tissue fixatives (like Bouin's solution, which contains picric acid) and treatments (like tissue decalcification) can harm DNA and render tissues useless for molecular analysis. These simple guidelines should be kept in mind when decisions are made regarding sources of DNA templates, especially when archived specimens are being considered for use in a study. Preparation of RNA typically requires fresh or frozen tissues, although techniques are being developed that promise to yield analyzable RNA from paraffin tissue sections.

3.2. DNA POLYMERASE ENZYMES IN PCR A DNA polymerase enzyme is required for DNA synthesis during the primer extension step of PCR. The contemporary PCR employs *Taq* DNA polymerase (isolated from *T. aquaticus*) *(11)*. *Taq* polymerase exhibits 5′→3′ polymerase activity, 5′→3′ exonuclease activity, thermostability, and optimum performance at 70–80°C *(5,12)*. Temperature, pH, and ion concentrations (Mg^{2+}) can influence the activity of *Taq* polymerase. The half-life of *Taq* activity at 95°C is approx 40–60 minutes *(13,14)*, and extremely high denaturation temperatures (>97°C) will significantly reduce its active lifetime. Because time at temperature represents the critical parameter for maintenance of *Taq* activity, lowering of the denaturation temperature or reduction in the denaturation time can prolong the activity of the enzyme during PCR. The optimum pH for a given PCR will be between 8.0 and 10.0, but it must be determined empirically. The typical PCR will be carried out in a buffer (usually Tris-Cl) of pH 8.3. *Taq* polymerase requires divalent cations in the form of Mg^{2+}. Lower divalent cation (Mg^{2+}) concentrations decrease the rate of dissociation of enzyme from the template by stabilizing the enzyme–nucleic acid interaction *(15)*. Most

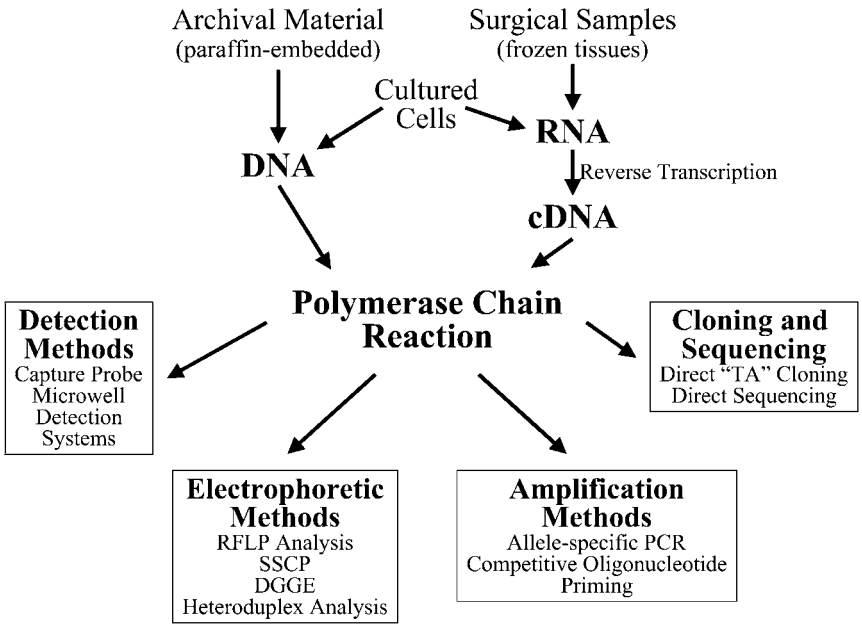


Fig. 2. Sources of nucleic acid templates for PCR and derived analytical methods. PCR analysis can be performed using DNA or cDNA (RNA), which can be prepared from a variety of sources, including those shown and others (such as forensic samples). PCR forms the basis numerous analytical techniques, some of which are based on differential amplification of target sequences. Other PCR-based methods involve advanced applications for analysis of PCR products (such as electrophoretic methods for mutation detection).

PCR mixtures will contain at least 1.5 m*M* $MgCl_2$. However, a $MgCl_2$ titration is recommended for any new template–primer combination.

Although *Taq* DNA polymerase is ideal for routine PCR, there are several other thermostable DNA polymerases with unique qualities *(16)* that make them useful for special PCR applications such as amplification of long pieces of DNA or high-fidelity amplification. *Taq* polymerase displays an error rate of approx 1×10^{-4} to 2×10^{-5} errors/bp (including both base substitutions and frameshift errors) *(17,18)*. Vent polymerase (also known as *Tli* polymerase from *Thermococcus litoralis*) exhibits an error rate of 2×10^{-5} to 5×10^{-5} errors/bp *(19,20)*. In addition to its increased fidelity over *Taq* polymerase, Vent polymerase has a half-life of approx 400 min at 95°C *(19)*, making it a very hardy enzyme. The highest-fidelity polymerase for PCR applications is *Pfu* polymerase (from *Pyrococcus furiosus*), which displays an error rate of 1.5×10^{-6} errors/bp *(21)*. The increased fidelity observed with Vent and *Pfu* polymerases (compared to *Taq*) is related to the 3′→5′ exonuclease proofreading activity that is inherent to these enzymes. *Taq* polymerase lacks 3′→5′ exonuclease proofreading activity. For high-fidelity applications, such as PCR amplification of DNA for sequence determination, the use of a higher-fidelity polymerase is preferred. Some manufacturers also produce blends of DNA polymerases for specialized PCR applications. One of these is referred to as Expand High Fidelity polymerase blend, which contains *Taq* (for polymerase activity) and *Pwo* (for proofreading). This blend is used in long PCR applications that require a higher fidelity than *Taq* can provide when used alone.

3.3. DESIGN OF OLIGODEOXYNUCLEOTIDE PRIMERS FOR PCR Design of oligonucleotide primers could be the most critical factor in determining the success of a PCR. Effective oligonucleotide primers for PCR are highly specific, free of secondary structure, and form stable duplexes with target sequences. Basically, four parameters need to be considered when designing a set of oligonucleotide primers: (1) size of the target sequence to be amplified, (2) the location of the target sequence within the overall genomic DNA (or cDNA) sequence, (3) secondary structure within the target and flanking regions, and (4) specificity of amplification. The size of the target sequence should be selected such that the PCR products produced range from 400 to 2000 bp in length. Products less than 400 bp in length are difficult to resolve using standard agarose gel electrophoresis techniques and might be obscured by excess primers or PCR artifacts. Products larger than 2000 bp might be amplified less efficiently because of the limited processivity of *Taq* polymerase *(22)*. Primer length can influence target specificity and efficiency of hybridization. A long oligonucleotide primer might be more specific for the target sequence, but it is less efficient at hybridization, whereas a short oligonucleotide primer is efficient at hybridization but less specific for the target sequence *(23)*. As a general guideline, oligonucleotide primers should be 17–30 nucleotides in length. Whenever possible, both primers should be of the same length because oligonucleotide primer length influences the calculated optimal annealing temperature for a specific primer. The base composition of the oligonucleotide primers is also important, because annealing temperature is governed in part by the G + C content of the primers. Ideally, G + C content should be 50–60%, and the percent G + C should be the same or very similar for both oligonucleotide primers in any given primer pair. The 3′-terminus of an oligonucleotide primer should contain a G, C, GC, or CG. Given the tighter hydrogen-bonding between G : C pairs, the presence of these nucleotides at the 3′ end of the oligonucleotide primer reduces the possibility for excessive breathing of the target–primer duplex, increasing the

efficiency of primer extension. However, runs of C or G at the 3′ end of an oligonucleotide primer should be avoided, as these can cause nonspecific hybridization with GC-rich sequences. Repetitive or palindromic sequences should be avoided in an oligonucleotide primer, and primer pairs should not contain sequences that are complementary to each other. Likewise, oligonucleotide primer pairs should not hybridize elsewhere in the gene being amplified or in other sequences contained within the genome. Web-based tools are available for easily analyzing the characteristics and properties of selected oligonucleotide primers (http://www.basic.nwu.edu/biotools/oligocalc.html).

When designing oligonucleotide primers for reverse transcriptase (RT)-PCR applications, all of the general guidelines for oligonucleotide primer design apply, but a few additional special considerations are needed. The precise location of the target sequence within the gene of interest must be chosen carefully. This is related to the fact that reverse transcription of mRNA templates using oligo(dT) as a primer often fails to generate full-length cDNA transcripts. Thus, amplification of target sequences that are distant from the 3′ end of the mRNA might be difficult to accomplish. Selection of target sequences that are within 1 kb of the 3′-terminus of the mRNA will ensure successful PCR amplification in most cases. However, if amplification of target sequences in the 5′ region of a gene is required, there are specialized techniques that can facilitate this amplification. One option is to use a gene-specific oligonucleotide primer during the reverse-transcription step, which will give rise to a cDNA template that includes the target sequence of interest. Another method for obtaining 5′-terminal sequence from mRNA is referred to as 5′ RACE (for rapid amplification of cDNA ends) *(24)*. RACE techniques can be applied to either the 3′ or 5′ ends of the mRNA transcript and are particularly useful when the mRNA sequence in that region of the gene is not known. mRNA secondary structure is another reason for the failure of RT to synthesize full-length cDNA transcripts, so regions known to have bulky secondary structure should be avoided if possible. RNA folding algorithms capable of predicting secondary structure within regions of known sequence are available *(25)*.

The optimal annealing temperature (T_m) for a given oligonucleotide primer set is very important for correctly setting up an effective PCR. The melting temperature of an oligonucleotide primer is most accurately calculated using nearest-neighbor thermodynamic calculations represented in the following formula: $T_m = \mathrm{H}[\mathrm{S} + R\ln(c/4)] - 273°\mathrm{C} + 16.6\log_{10}[\mathrm{K}^+]$, where H is the enthalpy and S is the entropy for helix formation, R is the molar gas constant, and c is the concentration of primer. Using such a formula is most easily accomplished using software for primer design and analysis, including some Web-based programs (such as http://www.rnature.com/oligonucleotide.html or http://alces.med.umn.edu/rawtm.html). However, an excellent working approximation for primer melting temperature can be calculated using simplified formulas that are generally valid for oligonucleotide primers that are 18–24 bp in length. One such simple formula for calculating annealing temperature for any given primer is $T_m = 69.3 + 0.41(\%\mathrm{G} + \mathrm{C}) - (650/L)$, where L is the primer length in bases *(26)*. Another formula for this calculation is $T_m = 2(\mathrm{A} + \mathrm{T}) + 4(\mathrm{G} + \mathrm{C})$, where A + T and G + C refer to the number of bases in each group. Primer annealing temperatures between 55°C and 72°C are preferred, but many standard PCR primers will have annealing temperatures of 55–65°C.

Several computer algorithms and programs have been developed to facilitate the design of primers with appropriate characteristics for PCR applications. Numerous Web-based oligonucleotide primer design programs are available, including OligoPerfect Designer (http://www.invitrogen.com), Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi), and DoPrimer (http://doprimer.interactiva.de/index.html). Web-based tools are also available for designing primers that are specific for exonic and intronic sequences of genes of interest (http://ihg.gsf.de/ihg/ExonPrimer.html) and for making degenerate oligonucleotide primers (http://bioinformatics.weizmann.ac.il/blocks/codehop.html).

3.4. REACTION MIXTURES FOR PCR A typical PCR mixture will include a reaction buffer, oligonucleotide primers, *Taq* polymerase, and an appropriate DNA template. The PCR buffer consists of 50 m*M* KCl, 1.5 m*M* $MgCl_2$, 10–50 m*M* Tris-Cl (pH 8.3), and 50–200 μ*M* dNTPs. Concentrations of KCl that are higher than 50 m*M* can inhibit *Taq* polymerase and should be avoided. However, the presence of KCl is necessary to encourage primer annealing to the template DNA. Likewise, excessive NaCl concentrations in a PCR mixture can adversely affect the activity of *Taq* polymerase. The amount of $MgCl_2$ that is optimal for a given PCR must be determined empirically. However, most standard PCR can be accomplished using 1.5–2 m*M* $MgCl_2$. The final concentration of dNTPs is 200 μ*M* for a typical PCR, but some applications can be accomplished using much lower concentrations. Higher concentrations of dNTPs (or $MgCl_2$) can encourage errors related to dNTP misincorporation by *Taq* polymerase and should be avoided. The typical PCR mixture will include 0.2–1 μ*M* of each oligonucleotide primer. The concentration of primers should not exceed 1 μ*M* unless the primers employed contain a high degree of degeneracy. *Taq* polymerase is provided from the supplier at 5 U/μL. One unit of enzyme activity is defined as the amount of enzyme required to catalyze the incorporation of 10 nmol of dNTP into acid-insoluble material in 30 min under standard reaction conditions. The amount of *Taq* polymerase included in a PCR will depend on the reaction size (20–50 μL). A 50-μL reaction will typically require 2.5 U of enzyme activity. The amount of DNA template included in a PCR will vary with the nature of the template source and the target sequence. Amplification from genomic DNA might require as much as 100 ng of DNA for a 50-μL reaction, whereas amplification from a plasmid template might only require 5 ng of DNA. Likewise, amplification of a target sequence that corresponds to a single allele might require more template, whereas amplification of a repetitive sequence (like Alu) will require substantially less template.

The various components of the PCR mixture can be prepared in the laboratory or can be purchased from commercial sources. A prepared reaction buffer can be obtained for use with *Taq* polymerase (Perkin-Elmer Cetus or other suppliers of the enzyme) and other thermostable polymerase enzymes. Likewise, commercially prepared dNTP stock solutions can be purchased from several sources. Complete PCR mix-in-a-tube reaction mixtures can be purchased that contain all of the

required PCR components except DNA template, oligonucleotide primers, and *Taq* polymerase (such as the EasyStart Micro 50 PCR mix-in-a-tube from Molecular BioProducts, http://www.mbpinc.com/). In fact, some commercially prepared PCR mixtures are supplied in a form that includes the *Taq* polymerase (such as PCR Master Mix from Promega, http://www.promega.com), requiring the addition of only oligonucleotide primers and DNA template. These commercially prepared reaction mixtures are extremely consistent and reliable and they work for many routine PCR applications.

3.5. REACTION MIXTURE ADDITIVES FOR PCR The inclusion of gelatin or bovine serum albumin (BSA), which could be included at concentrations up to 100 µg/mL, can enhance the efficiency of PCR. These agents act to stabilize the polymerase enzyme. The addition of helix destabilizing chemicals might be necessary if the target sequence for PCR is known to be of high G + C content *(27)*. For example, dimethyl sulfoxide (DMSO), dimethyl formamide (DMF), formamide, or urea could be included for this purpose. In most cases, these additives are included in the reaction mixture at 10% (w/v or v/v). These additives are thought to lower the T_m of the target sequence. Care must be used when these additives are incorporated into a PCR because high concentrations of these chemicals can adversely affect polymerase activity. For instance, concentrations of DMSO that exceed 10% can decrease *Taq* polymerase activity by as much as 50%. Many PCR mixtures will include nonionic detergents such as Tween-20, Triton X-100, or Nonident P40 at 0.05–0.1%. The inclusion of these detergents can increase enzyme activity for certain polymerase preparations by preventing enzyme aggregation.

4. OPTIMIZATION OF PCR REACTIONS

A number of different factors can significantly impact on PCR sensitivity and specificity, including (1) oligonucleotide primer design, (2) PCR cycling parameters (number of cycles, cycle times, and temperatures), and (3) the composition of the PCR mixture (Mg^{2+} concentration). For most PCR applications, the most critical parameter is the annealing temperature for the oligonucleotide primers employed. The maximum annealing temperature is determined by the primer with the lowest T_m. Exceeding this T_m by more than several degrees will diminish the ability of the oligonucleotide primer to anneal to the target sequence and could result in the failure to produce the product of interest. If an annealing temperature equal to T_m of the oligonucleotide primers fails to produce the desired product, then it might be necessary to lower the annealing temperature further. If the desired amplicon is produced but the level of background products is high, then the annealing temperature should be increased. Salt concentrations also affect several aspects of PCR. Mg^{2+} concentration can affect oligonucleotide primer annealing to the target sequence, the T_m of oligonucleotide–template complexes, and enzyme activity and fidelity. *Taq* polymerase requires free Mg^{2+} for activity. Therefore, sufficient $MgCl_2$ must be included in the PCR mixture to provide sufficient Mg^{2+} for the enzyme after some of the cation is lost to chelation by the oligonucleotide primers and the template DNA. The concentration of other salts can affect the PCR as well (including KCl). However, optimization of most PCR applications can be achieved through modification of Mg^{2+} concentration.

Complete optimization of the reaction conditions might require several adjustments to the annealing temperature, PCR cycle parameters, and salt concentrations. PCR optimization kits are commercially available. These kits include various PCR buffers and enable modification of various components of the PCR mixture, enabling optimization in less time and with fewer steps. For instance, the PCR Optimization Kit (from Roche Applied Science, http://www.roche-applied-science.com) contains a set of 16 buffers facilitating the simultaneous variation of pH and $MgCl_2$ concentration, as well as PCR mixture additives for further optimization (DMSO, gelatin, and glycerol).

5. INCREASING PCR SPECIFICITY AND SENSITIVITY

Several approaches are available that can improve upon the specificity and sensitivity of various PCR applications. Some of these approaches utilize specialized reagents, while others employ modified PCR cycling parameters or variations of oligonucleotide primer design.

5.1. IMPROVED *TAQ* POLYMERASE PREPARATIONS Several specialized PCR reagents have emerged over the last decade that function to improve PCR specificity. These include special preparations of *Taq* polymerase. *Taq* polymerase has substantial polymerase activity at 37°C, even though its optimal activity is expressed at much higher temperatures (approx 72°C). This low-temperature polymerase activity is the basis for some mispriming events that occur during the initial phase of PCR. Extension can occur from oligonucleotide primers that anneal nonspecifically to template DNA prior to the first denaturation step at 93–95°C. Platinum *Taq* Polymerase (from Invitrogen Life Technologies, http://www.invitrogen.com) was introduced to combat this problem by including a thermolabile inhibitor of the *Taq* polymerase in the form of monoclonal antibodies. During the initial denaturation phase of PCR, the elevated temperature destroys the monoclonal antibodies, releasing the *Taq* polymerase to function in PCR. The antibody-mediated inhibition of *Taq* allows for room-temperature reaction assembly. *Taq*Bead Hot Start Polymerase is a related product from Promega (http://www.promega.com). In this preparation, the *Taq* polymerase is physically sequestered within wax beads. The *Taq* polymerase becomes mixed with the PCR mixture during the initial denaturation step of PCR when the wax bead melts (at approx 60°C). In both cases, the *Taq* polymerase is functionally inactive until a critical temperature is reached, eliminating or reducing the possibility of primer extension from mispriming events.

5.2. HOT-START PCR Hot-start PCR was developed to reduce background from nonspecific amplification by preventing polymerization of new DNA during the initial phase of the reaction, when nonspecific binding could occur between primers and other DNAs in the mixture *(28,29)*. Hot start could be achieved by limiting the initial concentration of one or more PCR mixture components, including $MgCl_2$ or dNTP. However, the more commonly applied version of hot-start PCR separates the components of the PCR mixture with a barrier that is removed when the PCR is initiated. Wax beads form the

basis for the typical physical barrier employed in this application. Critical components of the reaction, including *Taq* polymerase, are separated from the remainder of the reaction by the wax barrier, which melts as the mixture is heated. The key concept in hot-start PCR is to achieve a temperature greater than the annealing temperature of the oligonucleotide primers prior to the complete mixing of the reaction components initiating polymerization.

5.3. TOUCH-DOWN PCR Touch-down PCR was developed to enhance amplification of desired target sequences while reducing amplification from mispriming events or from other PCR artifacts *(30,31)*. In this approach, the initial PCR cycle begins with an annealing temperature that is greater than the calculated T_m of the oligonucleotide primers utilized. The elevated annealing temperature employed during early cycles will encourage high specificity of primer annealing, although at lower efficiency. In subsequent cycles, the annealing temperature is progressively and incrementally lowered, improving the efficiency of oligonucleotide primer annealing. By manipulating the annealing temperature over the course of PCR, the desired amplicon will preferentially accumulate with good yield while amplification of undesirable products is minimized.

5.4. IMPROVING PCR SENSITIVITY Polymerase chain reaction is very sensitive and detectable levels of amplified products can usually be produced in a standard reaction consisting of 25–30 cycles. However, given the complexities of the factors that determine yield of PCR products (including the starting concentration of target sequence within the template and the efficiency of the reaction), there will be times when a standard reaction will not yield sufficient product to be visualized or otherwise detected. There are several strategies to overcome this problem. The simplest solution is to perform a second round of PCR using a small quantity of the first-round reaction mixture as the template source. In most cases, 25–30 additional cycles of PCR using this target-enriched template will result in detectable product. A second solution to this problem is nested primer PCR. In this application, the first-round PCR is accomplished using one set of primers, and the second-round PCR is accomplished using oligonucleotide primers that are target-sequence-specific, but sited internal to (or nested within) the first oligonucleotide primer pair. Thus, the first-round of PCR provides a target-enriched template for the second-round of amplification using the nested primers. The advantage of nested PCR over a simple second-round PCR involves the specificity of the amplification reactions. By using distinct PCR primers that are directed to the same target sequence, nested PCR maintains a high level of specificity in the second round of amplification. Nonspecific products produced during the first round of PCR are not likely to contain sequences complementary to the nested primers. Thus, spurious first-round amplicons are eliminated during the second round of PCR. In contrast, second-round amplification using the same PCR primers is less stringent and will result in amplification of both the desired template as well as any nonspecific products from the first reaction. In addition, extremely rare target sequences can be detected using nested PCR. The improved sensitivity of nested PCR is related to the fact that the first round of PCR effectively enriches the template sample prior to the second round of PCR. Because of the sensitive nature of nested PCR, special care must be taken in order to avoid contamination.

6. PCR CONTAMINANTS

When performing PCR amplification, it is always important to be aware of potential sources of DNA contamination. The power to amplify very small amounts of DNA to detectable levels demands that special care be taken to prevent cross-contamination between different samples. This is especially true for PCR targets expected to be present in low numbers because greater efforts are usually required to amplify these sequences. Sources of contamination include (1) genomic DNA contaminating RNA samples, (2) cross-contamination among different nucleic acid samples processed simultaneously, (3) laboratory contamination of cloned target sequences (genomic or cDNA), and (4) carryover of PCR products. In general, the likelihood of contamination is substantially reduced by working in a clean laboratory and using good laboratory practices (wearing clean gloves at all times). Carryover products from other PCR reactions can be effectively controlled by the use of aerosol-free pipet tips, by using dedicated pipettors and solutions, and by maintaining separate areas to handle pre-PCR and post-PCR solutions and samples. In addition to these simple precautions, certain PCR strategies can reduce the possibility of drawing conclusions from false-positive results. For instance, in RT-PCR applications, the use of oligonucleotide primers that are positioned within different exons of the gene of interest will facilitate distinction of PCR products from the RNA template vs amplicons resulting from the amplification of contaminating genomic DNA. In this example, PCR products derived from genomic DNA will be of a different size (larger) than those from cDNA. In all PCR applications, it is essential to include proper positive and negative control reactions to guard against systematic contamination of PCR reagents and to ensure that the desired amplicon is produced in positive reactions.

7. ANALYSIS OF PCR PRODUCTS

There are numerous methods for analysis of PCR products (Figs. 2 and 3). The method of choice for analysis of PCR products will depend on the type of information that is desired. Typical analysis of PCR products will involve electrophoretic separation of amplicons and visualization with ethidium bromide or other DNA dye. In most cases, amplification products can be analyzed using standard agarose gel electrophoresis. Agarose gel electrophoresis effectively separates DNA products over a wide range of sizes (100 bp to >25 kbp). PCR products from 200 to 2000 bp can be separated quickly on a 1.6% agarose gel. When greater resolution or separation power is required, such as in the analysis of very small PCR products (<100 bp), polyacrylamide gel electrophoresis is the method of choice. In both cases, ethidium bromide can be added to the gel before casting or the gel can be stained with ethidium bromide following electrophoresis. DNA products are easily visualized by ultraviolet illumination after ethidium bromide staining. Another method often used to quantify products is the incorporation of radioactive, fluorescent, or biotinylated markers. PCR products can be labeled by incorporating labeled nucleotides or

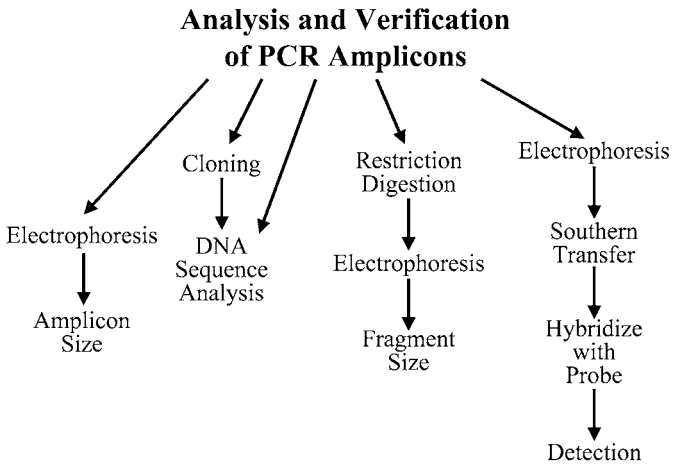


Fig. 3. Analysis and verification of PCR products. Amplicons produced using standard PCR can be analyzed and verified using various methods, some of which involve direct visualization of the PCR product (typically based on analysis of amplicon size). Other methods characterize the nature of the PCR product (based on sequence or sequence characteristics) or employ the PCR product in hybridization techniques.

through the use of labeled oligonucleotide primers. Labeled PCR products are separated by electrophoresis on either agarose or polyacrylamide gels and visualized using appropriate techniques (such as autoradiography for radioactively labeled products).

In some cases, the desired information resulting from a PCR analysis can be obtained through a simple analytical gel separation, whereas in other instances, additional information is required. PCR products can be cloned and used for sequence analysis, construction of molecular probes, mutation analysis, in vitro mutagenesis, studies of gene expression, and many other applications. PCR fragments can be introduced into suitable vectors via several methods *(15)*, where they can be expanded and/or manipulated. A cloning strategy should be planned prior to the PCR reaction because modifications of the product, such as the insertion of restriction sites, are sometimes necessary to allow insertion of the product into the intended vector. Care should also be taken to verify the size and purity of the PCR product prior to cloning. With careful planning and the inclusion of appropriate controls, the conventional techniques of recombinant DNA technology can be replaced almost entirely with PCR-based methods.

8. VERIFICATION OF PCR PRODUCTS

Although DNA bands of the expected size on an agarose or polyacrylamide gel are an encouraging sign that the reaction produced the desired amplicons, the identity of the PCR product should be verified by a secondary method. Verification is usually accomplished by nucleotide sequencing, restriction mapping, or sequence-specific probe hybridization (Fig. 3). Sequencing a PCR product provides the best information on the identity of the amplicon and represents the best method for verification of PCR products. Typically, the PCR product is cloned and sequenced by standard methods for either double-stranded or single-stranded DNA. With the advent of microsequencing techniques, purified products can be sequenced without having to be cloned *(32,33)*, shortening the time needed to verify a product considerably by eliminating the cloning step. Restriction mapping of PCR products is probably the simplest verification method. To use this method, the amplified sequence must contain one or more known unique restriction sites. Thus, products can be verified by digestion with the appropriate restriction endonuclease and subsequent separation and visualization by gel electrophoresis. Sequence-specific hybridization can also be used to verify PCR products. In this method, a synthetic oligonucleotide that recognizes a unique sequence within the amplified region is hybridized to the PCR product. Successful hybridization is measured by autoradiography (or related technique) and the resulting autoradiogram can be used to quantitate the amount of product produced. This method has the added advantage that stringent hybridization and washing conditions can be used along with other probes to differentiate between related gene transcripts that are similar in size.

9. INHIBITORS OF PCR

Organic and inorganic compounds that inhibit PCR amplification of nucleic acids are common contaminants in DNA samples from various origins. These contaminating substances can interfere with the PCR at several levels, leading to different degrees of attenuation and even to complete inhibition. This constitutes an important problem for both research and clinical laboratories. A wide variety of PCR inhibitors have been reported and they appear to be particularly abundant in complex samples such as animal fluids and samples containing high bacterial concentrations. Most of these contaminants (polysaccharides, urea, humic acids, hemoglobin) exhibit similar solubility to DNA. As a consequence, they are not completely removed when typical extraction protocols are used in the preparation of template DNA (such as detergent, protease, and phenol–chloroform treatments). Several methods have been developed to avoid these contaminating substances. Some of these methods are simple but result in the loss of non-negligible amounts of the original sample, whereas others are very specific methods directed against specific forms of contaminant and might require expensive materials.

10. PCR ANALYSIS OF RNA

10.1. ISOLATION OF RNA One of the most important factors in generating high-quality full-length cDNA is the purity and integrity of the total RNA used as starting material. Chaotropic agents such as guanidinium chloride and guanidinium isothiocyanate are capable of dissolving cellular structures and proteins, causing nucleoproteins to dissociate rapidly from nucleic acids and inactivating RNAse enzymes. Several procedures using chaotropic and reducing agents for the isolation of RNA from cells and tissues have been reported in the literature. One procedure commonly used to isolate intact RNA from all types of tissues, even those rich in RNAse enzymes, was first reported by Chirgwin et al. *(34)*. Improvements to this method that allow sufficient amounts of RNA to be isolated from a number of samples in several hours have been published. Using the method of Chomcznski and Sacchi *(35)*, it is possible to isolate sufficient RNA for reverse transcription from

as few as 1×10^3 cells. RNA isolation kits that employ these and other improvements are commercially available from many of the molecular biology companies. One of the most popular of these commercial reagents is TRIzol (from Invitrogen, http://www.invitrogen.com), which is based directly on the method of Chomcznski and Sacchi *(35)*. RNA can be further purified by oligo(dT) selection. This step is not typically necessary in RT-PCR applications but might help simplify subsequent cDNA synthesis and amplification steps by removing genomic DNA contamination that can amplify along with the target cDNA during PCR. Genomic DNA can also be removed from RNA preparations prior to cDNA synthesis by treatment with DNAse I.

10.2. AMPLIFICATION OF RNA Reverse transcriptase–PCR is an excellent method for analysis of RNA transcripts, especially for measuring low-abundance species or working with limited amounts of starting material. Classic blotting and solution hybridization assays require much more RNA for analysis and lack the speed and ease of technique afforded by PCR-based applications. RT-PCR couples the tremendous DNA amplification powers of the PCR with the ability of RT to reverse-transcribe small quantities of total RNA (1 ng or less) into cDNA. Using total RNA rather that poly(A) purified RNA reduces the possibility of losing specific (rare) mRNAs during the purification process and allows the use of very small quantities of starting material (cells or tissues). Other advantages of RT-PCR include versatility, sensitivity, rapid turnaround time, and the ability to compare multiple samples simultaneously.

Reverse transcriptase–PCR is basically a four-step process: (1) RNA isolation, (2) reverse transcription, (3) PCR amplification, and (4) PCR product analysis. RNA is isolated from cells or tissue and used as a template in a reverse-transcription reaction that produces cDNA, which serves as a template for the PCR reaction. Reverse transcriptase (retroviral RNA-directed DNA polymerase) is the enzyme used to catalyze cDNA synthesis. The RT reaction consists of five components: (1) cDNA synthesis primer, (2) an appropriate RT buffer, (3) dNTPs, (4) RNA template (total RNA or mRNA), and (5) RT enzyme. There are several commercially available RT enzyme preparations that can be used in standard RT-PCR applications. These include RT from Moloney murine leukemia virus (MMLV) and avian myeloblastosis virus (AMV). More recently, recombinant derivatives of these RT enzymes have become available that offer advantages over the native enzymes. SuperScript III Reverse Transcriptase (from Invitrogen, http://www.invitrogen.com) is a mutant form of MMLV RT with increased thermal stability (half-life at 50°C of 220 min) and reduced RNAse H activity. Advanced enzyme preparations like these produce the highest yields and confer high specificity when gene-specific primers are employed. However, conventional RT-PCR applications employ oligo(dT) as the primer for RT. This type of primer is designed to bind specifically to the poly(A) tail of the mRNA, although these primers can also anneal to long stretches of adenosine within mRNA sequences. To combat this problem, anchored oligo(dT) primers have been generated by adding a single G, A, or C to the 3′ end of the oligo(dT) *(36)*. The PCR aspect of RT-PCR is identical to that described for DNA applications, with the exception that the template is cDNA rather than genomic DNA.

11. CONCLUSION

In retrospect, the impact of the PCR has been enormous for basic research, translational research, and clinical diagnostic purposes. The ease of use and continued modifications by many investigators over the years propelled a rather simple chemical reaction to the forefront of diagnostic medicine. In the subsequent chapters, other technologies and clinical applications are presented, many of which were influenced by the traditional PCR.

REFERENCES

1. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., et al. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230:**1350–1354, 1985.
2. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239:** 487–491, 1988.
3. Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symp. Quant. Biol.* **51 (Pt. 1):**263–273, 1986.
4. Mullis, K. B. and Faloona, F. A. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.* **155:**335–350. 1987.
5. Lawyer, F. C., Stoffel, S., Saiki, R. K., Myambo, K., Drummond, R., and Gelfand, D. H. Isolation, characterization, and expression in *Escherichia coli* of the DNA polymerase gene from *Thermus aquaticus*. *J. Biol. Chem.* **264:**6427–6437, 1989.
6. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., et al., eds. *Short Protocols in Molecular Biology,* 5th ed., Wiley, New York, 2002.
7. O'Leary, J. J., Engels, K., and Dada, M. A. The polymerase chain reaction in pathology. *J. Clin. Pathol.* **50:**805–810, 1997.
8. Brinkmann, B. Overview of PCR-based systems in identity testing. *Methods Mol. Biol.* **98:**105–119, 1998.
9. Rogers, B. B. Application of the polymerase chain reaction to archival material. *Perspect. Pediatr. Pathol.* **16:**99–119, 1992.
10. Cano, R. J., Poinar, H. N., Pieniazek, N. J., Acra, A., and Poinar, G. O., Jr. Amplification and sequencing of DNA from a 120–135-million-year-old weevil. *Nature* **363:**536–538, 1993.
11. Vieille, C. and Zeikus, G. J. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* **65:**1–43, 2001.
12. Lawyer, F. C., Stoffel, S., Saiki, R. K., Chang, S. Y., Landre, P. A., Abramson, R. D., et al. High-level expression, purification, and enzymatic characterization of full-length *Thermus aquaticus* DNA polymerase and a truncated form deficient in 5′ to 3′ exonuclease activity. *PCR Methods Appl.* **2:**275–287, 1993.
13. Chien, A., Edgar, D. B., and Trela, J. M. Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *J. Bacteriol.* **127:**1550–1557, 1976.
14. Kaledin, A. S., Sliusarenko, A. G., and Gorodetskii, S. I. [Isolation and properties of DNA polymerase from extreme thermophylic bacteria *Thermus aquaticus* YT-1]. *Biokhimiia* **45:**644–651, 1980.
15. Fanning, S. and Gibbs, R. A. PCR in genome analysis, *in Genome Analysis. Volume 1: Analyzing DNA,* Birren, B., Green, E. D., Klapholz, S., Myers, R. M., and Roskams, J., eds., Cold Spring Harbor Laboratory Press, Plainview, NY, pp. 249–299, 1997.
16. Perler, F. B., Kumar, S., and Kong, H. Thermostable DNA polymerases. *Adv. Protein Chem.* **48:**377–435, 1996.
17. Eckert, K. A. and Kunkel, T. A. DNA polymerase fidelity and the polymerase chain reaction. *PCR Methods Appl.* **1:**17–24, 1991.
18. Tindall, K. R. and Kunkel, T. A. Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Biochemistry* **27:**6008–6013, 1988.

19. Mattila, P., Korpela, J., Tenkanen, T., and Pitkanen, K. Fidelity of DNA synthesis by the *Thermococcus litoralis* DNA polymerase—an extremely heat stable enzyme with proofreading activity. *Nucleic Acids Res.* **19:**4967–4973, 1991.
20. Cariello, N. F., Swenberg, J. A., and Skopek, T. R. Fidelity of *Thermococcus litoralis* DNA polymerase (Vent) in PCR determined by denaturing gradient gel electrophoresis. *Nucleic Acids Res.* **19:**4193–4198, 1991.
21. Lundberg, K. S., Shoemaker, D. D., Adams, M. W., Short, J. M., Sorge, J. A., and Mathur, E. J. High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus. Gene* **108:**1–6, 1991.
22. Bloch, W. A biochemical perspective of the polymerase chain reaction. *Biochemistry* **30:**2735–2747, 1991.
23. Wu, D. Y., Ugozzoli, L., Pal, B. K., Qian, J., and Wallace, R. B. The effect of temperature and oligonucleotide primer length on the specificity and efficiency of amplification by the polymerase chain reaction. *DNA Cell Biol.* **10:**233–238, 1991.
24. Frohman, M. A., Dush, M. K., and Martin, G. R. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85:**8998–9002, 1988.
25. Pallansch, L., Beswick, H., Talian, J., and Zelenka, P. Use of an RNA folding algorithm to choose regions for amplification by the polymerase chain reaction. *Anal. Biochem.* **185:**57–62, 1990.
26. Wallace, R. B., Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T., and Itakura, K. Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch. *Nucleic Acids Res.* **6:**3543–3557, 1979.
27. Sarkar, G., Kapelner, S., and Sommer, S. S. Formamide can dramatically improve the specificity of PCR. *Nucleic Acids Res.* **18:**7465, 1990.
28. Kaijalainen, S., Karhunen, P. J., Lalu, K., and Lindstrom, K. An alternative hot start technique for PCR in small volumes using beads of wax-embedded reaction components dried in trehalose. *Nucleic Acids Res.* **21:**2959–2960, 1993.
29. Bassam, B. J. and Caetano-Anolles, G. Automated "hot start" PCR using mineral oil and paraffin wax. *Biotechniques* **14:**30–34, 1993.
30. Roux, K. H. Using mismatched primer-template pairs in touchdown PCR. *Biotechniques* **16:**812–814, 1994.
31. Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K., and Mattick, J. S. 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* **19:**4008, 1991.
32. Bachmann, B., Luke, W., and Hunsmann, G. Improvement of PCR amplified DNA sequencing with the aid of detergents. *Nucleic Acids Res.* **18:**1309, 1990.
33. Casanova, J. L., Pannetier, C., Jaulin, C., and Kourilsky, P. Optimal conditions for directly sequencing double-stranded PCR products with sequenase. *Nucleic Acids Res.* **18:**4028, 1990.
34. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18:**5294–5299, 1979.
35. Chomczynski, P. and Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Anal. Biochem.* **162:**156–159, 1987.
36. Liang, P. and Pardee, A. B. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257:**967–971, 1992.

6 Bioinformatics

Computer-Based Approaches to Genetic Analysis

SHARON L. RICKETTS

1. INTRODUCTION

The application of computers, databases, and computational methods to the capture and interpretation of biological data defines the field of bioinformatics. This is an integrative discipline, combining computer science, mathematics, physics, and biology, and is essential for almost every aspect of data management in molecular biology. Numerous organisms have been sequenced, either completely or to near completion. The bacteria *Haemophilus influenzae* was the first genome to be sequenced completely *(1)*. This was followed by sequencing of the first eukaryotic genome, the budding yeast *Saccharomyces cerevisiae (2)*. Other organisms followed, including the nematode *Caenorhabditis elegans (3)*, the fruit fly *Drosophila melanogaster (4)*, and the mouse *Mus musculus (5)*. A working draft of the human genome was released in February 2001 *(6)*, with fine sequencing continuing. Bioinformatics has played a critical role in the sequencing of all these organisms.

Genome sequencing of multicellular organisms, as well as many bacterial and viral pathogens, is key to our understanding of the molecular basis of inherited, somatic, and infectious diseases. Sequencing itself, however, is not the end; we also need to know the function of the genes. The mapping of the human genome will impact biomedical research, diagnosis, vaccines, therapeutics, and preventative health care, ultimately allowing for better patient management. The future of bioinformatics is limitless and will continue to impact many fields, including molecular diagnostics.

2. THE HUMAN GENOME PROJECT

The Human Genome Project (HGP) began in 1990 as a joint effort by the US Department of Energy and the National Institutes of Health. The project originally was planned to last 15 yr; however, a working draft of the human genome sequence was published in February 2001 *(6)*. The goals of the HGP included the identification of the approx 30,000 genes in human DNA, sequence determination of the 3 billion chemical basepairs that make up human DNA, the storage of this information in databases, the improvement of tools for data analysis, the transfer of related technologies to the private sector, and assessment of the ethical, legal, and social issues that might arise from the project (www.ornl.gov).

3. BIOINFORMATICS TOOLS

3.1. SEQUENCE DATABASES Sequence analysis is perhaps the most widely recognized component of the HGP and bioinformatics. There are three major international databases of publicly available information on nucleotide sequences. GenBank (www.ncbi.nlm.nih.gov/Genbank/index.html) is maintained by the National Center for Biotechnology Information in the United States, EMBL (www.ebi.ac.uk/) is maintained by the European Bioinformatics Institute in the United Kingdom, and the DNA Data Bank of Japan (www.ddbj.nig.ac.jp/intro-e.html) is retained in Japan. The Human Genome Browser Gateway (www.genome.ucsc.edu/cgi-bin/hgGateway) is another useful genome sequence information site. The databases contain identical sequence information, however, each site formats the information differently. There are tools available for sequence and structure prediction for a submitted nucleotide sequence on each website.

3.2. MAP VIEWER Map Viewer (www.ncbi.nlm.nih.gov/mapview/), available through the National Center for Biotechnology Information, is a tool that allows for visualization of an organism's complete genome, integrated maps for each chromosome, and sequence data for a genomic region of interest. The data are presented graphically, displaying available sequence data, cytogenetic, genetic, physical, and radiation hybrid maps.

3.3. BASIC LOCAL ALIGNMENT SEARCH TOOL The BLAST (basic local alignment search tool; www.ncbi.nlm.nih.gov/BLAST/) search program is a well-known tool that is designed to identify all similar nucleotide and protein sequences. These alignment algorithms compare experimentally derived sequences to one or more databases of either nucleotides or amino acids. Homology is determined by a greater than 30% match and implies that sequences might be related by divergence from a common ancestor or share common functional aspects. BLAST 2.0 is the currently available version of the program. Because of its ability to accommodate for introns in a DNA sequence, it is also referred to as gapped BLAST.

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

3.4. EEPRESSED SEQUENCE TAGS DATABASE AND UNIGENE One aspect of the HGP is the generation of an expressed sequence tag (EST) database composed of large numbers of low-quality cDNA sequences. ESTs are small pieces of DNA sequence that are generated by sequencing the 3′ and/or 5′ ends of expressed genes. Although their quality is low, the volume of ESTs produced makes them a good source for identifying new gene sequences. EST sequences are deposited into dbEST (www.ncbi.nlm.nih.gov/projects/dbEST/) or UniGene (www.ncbi.nlm.nih.gov/UniGene/index.html). dbEST was created to organize, store, and provide access to the large volume of EST data that have accumulated. Redundancy in dbEST is common, because of many ESTs that match the same gene. UniGene automatically partitions GenBank mRNA and EST sequences into a nonredundant set of gene-oriented clusters composed of identical 3′ untranslated regions *(7)*. A variety of species are represented in the database: human, rat, mouse, cow, zebrafish, clawed frog, fruitfly, mosquito, wheat, rice, barley, maize, and cress.

4. MAPPING OF CANDIDATE DISEASE GENES

The Human Genome Project was initiated in the United States on October 1, 1990, with the goal of elucidating the genetic architecture of the human genome. In February 2001, a working draft of the human genome was published *(6)*, and fine sequencing of the genome continues. The sequencing of the human genome impacts greatly on the study of human disease. The identification, isolation, and characterization of the genes involved in normal biological processes as well as disease states should greatly increase as a result of a known gene sequence. The completion of the HGP will result in the accumulation of vast amounts of sequencing data to be used for the discovery of fundamental biological processes, isolation of candidate disease genes, and the diagnosis, treatment and prevention of human disease.

The classical approaches to identifying genes involved in human disease include functional and positional cloning. In functional cloning, the disease gene is usually isolated only after the underlying physiological defect has been elucidated. The study of the gene function precedes gene identification and chromosomal mapping. Examples of disease genes isolated through functional cloning include phenylalanine hydroxylase in phenylketonuria *(8)* and the β-hemoglobin gene in β-thalassemia *(9)*. In positional cloning, gene mapping precedes the identification of the disease gene *(10)*. This approach is classically used when there is little or no understanding about the function of a defective gene. Generally, blood samples from families in which the genetic disorder is segregating are collected. Genetic mapping then limits the region of the genome to a size that allows molecular cloning of the interval, analysis of encoded genes, and identification of mutations. Finally, functional analysis can be applied to the gene. The mutated gene that gives rise to Duchenne muscular dystrophy was one of the many disease genes isolated via its chromosomal location by positional cloning *(11)*.

Our laboratory employs the products of the HGP and the technique of positional cloning to identify human liver tumor suppressor genes. The 11p11.2-p12 chromosomal region was demonstrated to harbor a candidate disease gene [in hepatocellular carcinoma (HCC)], not through pedigree analysis, as in classical positional cloning, but through the use of human/rat microcell hybrid cell lines. Analysis of chromosome 11 microsatellite markers indicated that the locus conferring tumor suppression in our model system was located on human 11p11.2-p12. More refined mapping was performed using EST markers that were in the candidate region *(12)*. All ESTs and genes identified within the delimited region of DNA (11p11.2-p12) became candidates for the disease (HCC). Through deletion mapping of suppressed MCH cell lines in our laboratory with EST markers, we were able to identify candidate liver tumor suppressor genes. EST sequence can then be compared to other sequence data for identification of homology to any other DNA sequence. The gene corresponding to the EST can then be identified through further mapping and sequence analysis. Characterization of candidate genes is then necessary to demonstrate that the disease is associated with a mutation in a particular candidate gene.

The HGP is evolving from mapped chromosomes to the determination of complete DNA sequence through the incorporation of partial genomic sequence tracts into the underlying architecture of previously mapped regions. Although the HGP has nearly finished the goal of completely sequencing the human genome, there are still gaps and long tracts of unfinished sequencing data. The influx of new sequence information is dramatic; the position of genes and markers in the draft sequence changes daily. It is important to note, however, that until this information is complete, researchers need to incorporate the new sequence data with mapped data they have generated in their own laboratories.

5. IDENTIFICATION OF CANDIDATE GENES

The HGP is rapidly and accurately generating genomic information that aids in the identification of unknown genes and disease loci. Once a disease-linked region is found, it can be scrutinized to identify target genes. Positional cloning is used to map the location of a human disease gene through marker analysis and to employ the mapped location on the chromosome to clone the gene and identify its function. Dong et al. *(13)* identified *KAI1* through positional cloning by complementing a defect in rat prostatic tumor cell lines following introduction of human chromosome 11. We have employed a similar strategy to identify human liver tumor suppressor genes by introducing human chromosome 11 into rat liver epithelial tumor cell lines. It is possible that DNA from the introduced chromosome could be lost from microcell hybrid cell lines when the technique of microcell-mediated chromosome transfer is used. Therefore, extensive analysis of each hybrid cell line is required to confirm the presence of the transferred chromosome. In human–human hybrids, polymorphic markers are used to distinguish endogenous DNA from the introduced allele. Tumor suppressor genes can be localized if a correlation between suppression of the tumorigenic phenotype and the presence of specific chromosomal regions is observed *(14–16)*. Through microsatellite polymerase chain reaction (PCR) analysis of specific chromosomal sequences, regions of chromosomal loss or retention can be distinguished rapidly. The application of this technology in human tumor models is hindered by the requirement for heterozygosity at the marker loci. There is no

such limitation in rodent–human hybrids, where the introduced human allele can be easily distinguished from sequences in the recipient rat cell genome.

Studies in our laboratory suggest the presence of a human liver tumor suppressor gene in the p11.2-p12 region of human chromosome 11 *(17–19)*. Introduction of human chromosome 11 into rat liver epithelial tumor cell lines via microcell-mediated chromosome transfer produced microcell hybrid (MCH) cell lines that exhibit suppression of tumorigenicity in vivo *(17)*. Because of the fragmentation of the chromosome when inserted into the donor cell lines, it was possible to define the region of the chromosome that conferred suppression on 11p11.2-p12 *(17)*. Further molecular characterization of the MCH cell lines localized the putative liver tumor suppressor locus to a 950-kb region of 11p11.2.-p12 *(18)* and showed that this same locus suppresses the tumorigenic potential of other rat liver tumor cell lines *(19)*. These studies demonstrated that a genetic locus (or loci) on human chromosome 11 could suppress the tumorigenic potential and alter the phenotypic properties of rat liver tumor cell lines. The information acquired in these previous studies made it possible to pursue advanced mapping of the 11p11.2-p12 region and placement of genes into the critical tumor suppressor region.

Human chromosome 11 contains several known or putative tumor suppressors whose deletion is critical to the pathogenesis of various malignancies. We have used a positional cloning strategy to identify candidate liver tumor suppressor genes from human chromosome region 11p11.2-p12 through investigation of a rat liver epithelial tumor cell line and derived MCH cell lines. We previously established a sequence tagged site (STS)-based map that defined the liver tumor suppressor locus to the D11S1361–D11S1357 interval of 11p11.2-p12 *(18)*. We used EST and gene markers from the HGP to map the p11.2-p12 region of chromosome 11 that corresponded to the critical region of tumor suppression in our MCH cell lines. We integrated genomic maps from GeneMap'99 GB4 panel, NCBI Map Viewer, and the Human Genome Browser (University of California–Santa Cruz) to obtain the order of EST markers in the 11p11.2-p12 region. The superimposition of this EST-based map on our STS-based map resulted in the identification of candidate liver tumor suppressor genes from 11p11.2-p12, as well as increasing the number of informative genetic markers corresponding to this chromosomal region. From the 11p11.2-p12 region, 142 EST and gene markers were analyzed through PCR-based deletion mapping. Of these markers, 19 mapped to the critical region of tumor suppression in our MCH cell lines. Eleven of the 19 were expressed in the MCH cell lines, but not a tumor cell line, suggesting that 1 of these might be involved in liver tumor suppression in this model system.

Refinement of the number of candidate liver tumor suppressor genes was accomplished by screening MCH-derived tumor cell lines and additional MCH cell lines for the presence and expression of these candidate genes. Retention of EST markers in the DNA MCH-derived tumor cell lines and/or continued expression of these genes, would argue against their role as a tumor suppressor. These "functional" gene-expression-based analyses narrowed the number of candidate liver tumor suppressor genes to three. During our investigation of the 11p11.2-p12 chromosomal region, the HGP laboratories were in the process of sequencing the human genome. The accumulation of more and more sequencing data into the GenBank database resulted in the positioning of known genes to 11p11.2-p12. We used these genes to continue the mapping of 11p11.2-12. Thirty-nine known genes were localized to the p11.2-p12 region of human chromosome 11. Screening of these genes with the tumor cell line and the MCH cell lines resulted in the identification of three candidate genes *(20)*. Investigations into the sequences of these candidates revealed that the three ESTs each shared 100% homology with three of the four newly identified genes, leaving the final number of candidate liver tumor suppressor genes at four *(21)*.

6. CLINICAL APPLICATION OF BIOINFORMATICS

6.1. IDENTIFICATION OF DISEASE GENES The isolation of human disease genes has important implications not only for understanding the molecular and cellular basis of the disease but also for prevention and treatment. Cloning the gene responsible for an inherited disease leads to the potential for developing molecular diagnostic tests for mutant alleles of the gene, such as the detection of mutations by PCR. A test can be developed that can identify individuals at risk for the disease and enable appropriate preventive measures. In addition, the ability to detect mutant alleles of disease genes opens the possibility of prenatal diagnosis to prevent transmission of the disease to future generations. Bioinformatics will play an important role in the identification of susceptibility genes as well as in helping provide insight into the molecular pathogenesis of disease, providing an opportunity for the development of targeted therapy *(22)*.

6.2. PERSONALIZED MEDICINE Personalized medicine seeks to understand why the rate of progression for a given disease, and the response to individual drugs, is unique among patients with the same disease *(23)*. A goal of personalized medicine is to treat the patient with the most appropriate therapy. Molecular diagnostics will enter into many aspects of personalized medicine. Early detection of disease, selection of appropriate treatment (i.e., tailored drug therapy), therapeutics, and monitoring of therapy will all be possible by combining genetic and clinical information with bioinformatics *(24)*.

6.3. PHARMACOGENETICS Pharmacogenetics is the study of the influence of genetic factors on the actions of drugs. An individual's response to a drug is the interaction between genetic and nongenetic factors. Genetic variants in the drug itself, disease pathway genes, and drug metabolizing enzymes are predictors of drug efficacy or toxicity *(25)*. Single-nucleotide polymorphism (SNP) genotyping will play a large role in predicting adverse drug reactions among individuals. SNPs are small stretches of DNA with only a 1-base difference and serve to distinguish the genetic material of one person from another *(26)*. There are over 1.4 million SNPs in the human genome, with over 60,000 located in coding regions *(27)*. Many SNPs have already been associated with changes in the metabolism or effects of commonly used drugs and are used in the clinical laboratory as molecular diagnostics *(28)*. SNPs are ideal elements for analysis in the molecular diagnostics laboratory because they help in understanding the genetic basis of human diseases. SNP mapping data from clinical trials can be used to

determine a common set of polymorphisms shared by patients who do not respond, or have an adverse reaction to, a particular drug. The detection of SNPs requires the use of bioinformatics for cataloging and analyzing information.

6.4. PHARMACOGENOMICS Pharmacogenomics is a discipline that aims to explain the inherited basis for differences in drug response between individuals *(28)*. It is defined as the application of whole-genome technologies for the prediction of the sensitivity or resistance of an individual's disease to drug therapy *(29,30)*. There is great potential for pharmacogenomics to yield important new molecular diagnostics that will become routine clinical laboratory tests, thereby allowing physicians and pharmacists to select drugs for individual patients.

6.5. GENETIC DATABASES Genetic databases are a stored collection of genetic samples in the form of blood or tissue that can be linked with medical, genealogical, or lifestyle information from a specific population, gathered by using a process of generalized consent *(31)*. Iceland is the first country to compile individualized genetic information on distinct populations, but numerous other countries are proposing to do so. The UK Biobank project is proposed to be "the world's biggest study of the role of nature and nurture in health and disease" (http://www.biobank.ac.uk/). UK Biobank will collect DNA samples, medical records, and lifestyle information of 500,000 people between 45 and 69 yr old. It will follow the participants' health status for more than 10 yr. The Biobank will provide researchers chances to correlate genetic traits with common diseases. The aim of genetic databases is to map genes for common diseases to improve the health of the populations involved. The understanding of genetic disease susceptibility will be phenomenal with the kind of information available in these databases. However, the ethical, legal, and social implications need to be considered very carefully before the project is undertaken.

7. BIOINFORMATICS AND DRUG DISCOVERY

Bioinformatics can be used by pharmaceutical companies to advance therapeutics in developing drugs for specific targets such as proteases, kinases, nuclear hormone receptors, 7-transmembrane proteins, chemokines, cytokines, and adhesion molecules *(32)*. Imatinib mesylate (Gleevec) is used successfully in the treatment of chronic myelogenous leukemia by interfering with an abnormal BCR-ABL tyrosine kinase protein *(33)*. Gleevec was identified by using a high-throughput screen for tyrosine kinase inhibitors while optimizing its activity for specific kinases. Bioinformatics played a key role in the discovery of this drug through the ability to identify and target specific genetic markers. The success of this drug suggests that therapies based on genomic information will be effective.

8. CONCLUSIONS

The role of bioinformatics is firmly ensconced in the molecular biology laboratory. Molecular diagnostics will facilitate the detection and characterization of disease, monitor patient response to drug therapy, and assist in the identification of genetic modifiers and disease susceptibility.

Analysis of the draft human genomic sequence has already led to the identification of numerous disease genes. The draft sequence has also been used to identify an enormous number of SNPs. These discoveries, as well as future discoveries, will have a deep impact on the future conduct of biomedical research. In the coming years, physicians will be able to rapidly diagnose existing genetic diseases, predetermine genetic risk for developing a disease, and prescribe medical intervention based on a person's genetic information, reducing the chance of an allergic, or fatal, drug reaction.

REFERENCES

1. Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult C. J., et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269:**496–512, 1995.
2. Goffeau, A. The yeast genome directory. *Nature* **387:**5, 1997.
3. *C. elegans* sequencing consortium. Genome sequence of the nematode *C. elegans***:** a platform for investigating biology. *Science* **282:**2012–2018, 1998.
4. Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., et al. The genome sequence of *Drosophila melanogaster*. *Science* **287:**2185–2195, 2000.
5. Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420:**520–562, 2002.
6. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., et al.. Initial sequencing and analysis of the human genome. *Nature* **409:**860–921, 2001.
7. Wheeler, D. L., Church, D. M., Lash, A. E., Leipe, D. D., Madden, T. L., Pontius, J. U., Schuler, G. D., et al. Database resources of the National Center for Biotechnology Information: 2002 update. *Nucleic Acids Res*. **30:**13–16, 2002.
8. Robson, K. J., Chandra, T., MacGillivray, R. T., and Woo, S. L. Polysome immunoprecipitation of phenylalanine hydroxylase mRNA from rat liver and cloning of its cDNA. *Proc. Natl. Acad. Sci. USA* **79:**4701–4705, 1982.
9. Orkin, S. H. and Nathan, D. G. The molecular genetics of thalassemia, in *Advances in Human Genetics. Volume 11*. Harris, H. and Hirschhorn, K., eds., Plenum, New York, p. 233, 1981.
10. Green, E. D., Cox, D. R., and Myers, R. M. The human genome project and its impact on the study of human disease, in *The Genetic Basis of Human Cancer.* Vogelstein, B., Kinzler, K. W., eds., McGraw-Hill, New York, pp. 33–63, 1998.
11. Monaco, A. P., Bertelson, C. J., Middlesworth, W., et al. Detection of deletions spanning the Duchenne muscular dystrophy locus using a tightly linked DNA segment. *Nature* **316:**842–845, 1985.
12. Boguski, M. S., Tolstoshev, C. M., and Bassett, D. E., Jr. Gene discovery in dbEST. *Science* **265:**1993–1994, 1994.
13. Dong, J. T., Lamb, P. W., Rinker-Schaeffer, C. W., et al. KAI1, a metastasis suppressor gene for prostate cancer on human chromosome 11p11.2. *Science* **268:**884–886, 1995.
14. Negrini, M., Sabbioni, S., Possati, L., et al. Suppression of tumorigenicity of breast cancer cells by microcell-mediated chromosome transfer: studies on chromosomes 6 and 11. *Cancer Res.* **54:**1331– 1336, 1994.
15. Reid, L. H., West, A., Gioeli, D. G., Phillips K. K., Kelleher, K. F., Araujo, D., Stanbridge, E. J., et al. Localization of a tumor suppressor gene in 11p15.5 using the G401 Wilms' tumor assay. *Hum. Mol. Genet.* **5:**239–247, 1996.
16. Reid, L. H., Crider-Miller, S. J., West, A., Lee, M. H., Massague, J., and Weissman, B. E. Genomic organization of the human p57KIP2 gene and its analysis in the G401 Wilms' tumor assay. *Cancer Res.* **56:**1214–1218, 1996.
17. Coleman, W. B., McCullough, K. D., Esch, G. L., et al. Suppression of the tumorigenic phenotype of a rat liver epithelial tumor cell line by the p11.2-p12 region of human chromosome 11. *Mol. Carcinog.* **13:**220–232, 1995.
18. Coleman, W. B., Esch, G. L., Borchert, K. M., et al. Localization of a putative liver tumor suppressor locus to a 950-kb region of human 11p11.2-p12 using rat liver tumor microcell hybrid cell lines. *Mol. Carcinog.* **19:**267–272, 1997.

19. Mahon, M. C., Driscoll, M. P., Glover, W. J., et al. Suppression of tumorigenicity of rat liver epithelial tumor cell lines by a putative human 11p11.2-p12 liver tumor suppressor locus. *Int. J. Oncol.* **14:**337–346, 1999.
20. Ricketts, S. L., Carter, J. C., and Coleman, W. B. Identification of three 11p11.2 candidate liver tumor suppressors through analysis of known human genes. *Mol. Carcinog.* **36:**90–99, 2003.
21. Jahn, J. E., Ricketts, S. L., and Coleman, W. B. Identification of candidate liver tumor suppressor genes from human 11p11.2 by transcription mapping of microcell hybrid cell lines. *Int. J. Oncol.* **22:**1303–1310, 2003.
22. Bayat, A. Science, medicine, and the future: bioinformatics. *Br. Med. J.* **324:**1018–1022, 2002.
23. Ginsburg, G. S. and McCarthy, J. J. Personalized medicine: revolutionizing drug discovery and patient care. *Trends Biotechnol.* **19:**491–496, 2001.
24. Jain, K. K. Personalized medicine. *Curr. Opin. Mol. Ther.* **4:**548–558, 2002.
25. Ross, J. S. and Ginsburg, G. S. The integration of molecular diagnostics with therapeutics. Implications for drug development and pathology practice. *Am. J. Clin. Pathol.* **119:**26–36, 2003.
26. Brookes, A. J. The essence of SNPs. *Gene* **234:**177–186, 1999.
27. Sachidanandam, R., Weissman, D., Schmidt, S. C., et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409:**928–933, 2001.
28. Evans, W. E. Pharmacogenomics: marshalling the human genome to individualise drug therapy. *Gut* **52(Suppl. 2):**10–18, 2003.
29. Rusnak, J. M., Kisabeth, R. M., Herbert, D. P, and McNeil, D. M. Pharmacogenomics: a clinician's primer on emerging technologies for improved patient care. *Mayo Clin. Proc.* **76:**299–309, 2001.
30. Hess, P. and Cooper, D. Impact of pharmacogenomics on the clinical laboratory. *Mol. Diagn.* **4:**289–298, 1999.
31. Austin, M. A., Harding, S., and McElroy, C. Genebanks: a comparison of eight proposed international genetic databases. *Community Genet.* **6:**37–45, 2003.
32. Elkin, P. L. Primer on medical genomics part V: bioinformatics. *Mayo Clin. Proc.* **78:**57–64, 2003.
33. Druker, B. J., Sawyers, C. L., Kantarjian, H., et al. Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome. *N. Engl. J. Med.* **344:** 1038–1042, 2001.

III MOLECULAR DIAGNOSTIC TECHNOLOGIES

7 PCR-Based Methods for Mutation Detection

IAN M. FRAYLING, EMMA MONK, AND RACHEL BUTLER

1. INTRODUCTION

Although many, even most, methods of mutation detection depend on polymerase chain reaction (PCR), in the majority of techniques PCR itself does not detect the actual mutation. Rather, PCR generates an amplicon that is then analyzed by some other method to find possible mutations within the amplicon, such as conformation-based techniques like single-stranded conformational polymorphism (SSCP) analysis, denaturing gradient gel electrophoresis (DGGE), or sequencing. However, there are some methods in which a modified PCR acts as the primary mutation-detection system, although some type of electrophoresis might be needed to separate the subsequent amplicons. These include real-time PCR, the amplification refractory mutation system (ARMS), quantitative fluorescent PCR (QF-PCR), or a derivative of the oligoligation assay, multiplex ligation-dependent probe amplification (MLPA). Also discussed is the single-nucleotide primer extension assay and a proprietary derivative of it called Pronto™ . Because it uses a DNA polymerase in a post-PCR extension step, it can be deemed to fall into the group of PCR-based methods of mutation detection. A primary limitation of these methods is that, with a few exceptions, they are only suitable for testing for mutations that have been previously detected and characterized by other techniques.

2. REAL-TIME PCR

Real-time PCR is the only purely PCR-based method of mutation detection, in that it does not require any adjunctive technique, such as electrophoresis. As is described in this section, the PCR is directly monitored within the reaction tube and the PCR primers themselves are designed to be specific for a particular sequence or sequences. Thus, real-time PCR is only of use in detecting previously described mutations found by some other technique.

In real-time PCR, the exponential phase of PCR is monitored as it occurs, using fluorescently labeled molecules *(1)*. During the exponential phase, the amount of PCR amplicon DNA present in the reaction tube is directly proportional to the amount of starting material specific to the PCR primer pair (or target sequence). Thus, the amount of emitted fluorescence is directly proportional to the amount of amplicon, which, in turn, is proportional to the starting amount of target sequence *(2,3)*. If required, this can then be used to measure the target copy number. Real-time PCR needs to be distinguished from reverse transcription–PCR (RT-PCR), although, for example, RT-PCR can be used in conjunction with real-time PCR to measure gene expression as well as test for mutations *(4,5)*. There are two types of real-time PCR: that which uses nonspecific DNA binding dyes such as ethidium bromide or SYBR green I and that which uses labeled probes. These are discussed in the following subsections.

2.1. REAL–TIME PCR: NONSPECIFIC DNA-BINDING DYES Nonspecific DNA binding dyes simply allow the determination of the presence or absence of an amplicon, without giving any information regarding the precise nature of the product. SYBR Green I is a dye that emits fluorescence when it is bound to double-stranded DNA. As the PCR proceeds and the copy number of the product increases, the amount of intercalated SYBR Green I will also increase, raising the level of emitted fluorescence in direct proportion to the copy number *(6)*. Most applications of real-time PCR have been to measure gene expression or detect pathogens, although there are applications specific to medical genetics *(7–10)*.

2.2. REAL–TIME PCR: LABELED PROBES There are three main kinds of labeled probe for use in real-time PCR: Cleavage (5′ exonuclease) based, molecular beacons, and FRET probes (Fig. 1).

2.2.1. Cleavage-Based Probes Cleavage-based probes are the most widely used real-time PCR probes and depend upon the 5′ to 3′ exonuclease activity of *Taq* DNA polymerase (Fig. 1A). This is also known as the TaqMan® assay. Forward and reverse primers are bound to the target DNA sequence, and a fluorescently labeled probe specific for the wild-type or mutant sequence is bound downstream of the forward primer. The probe is an oligonucleotide that has a fluorescent reporter dye at the 5′ end and a quencher, typically TAMRA, attached to the 3′ end. It is modified to prevent extension occurring from

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

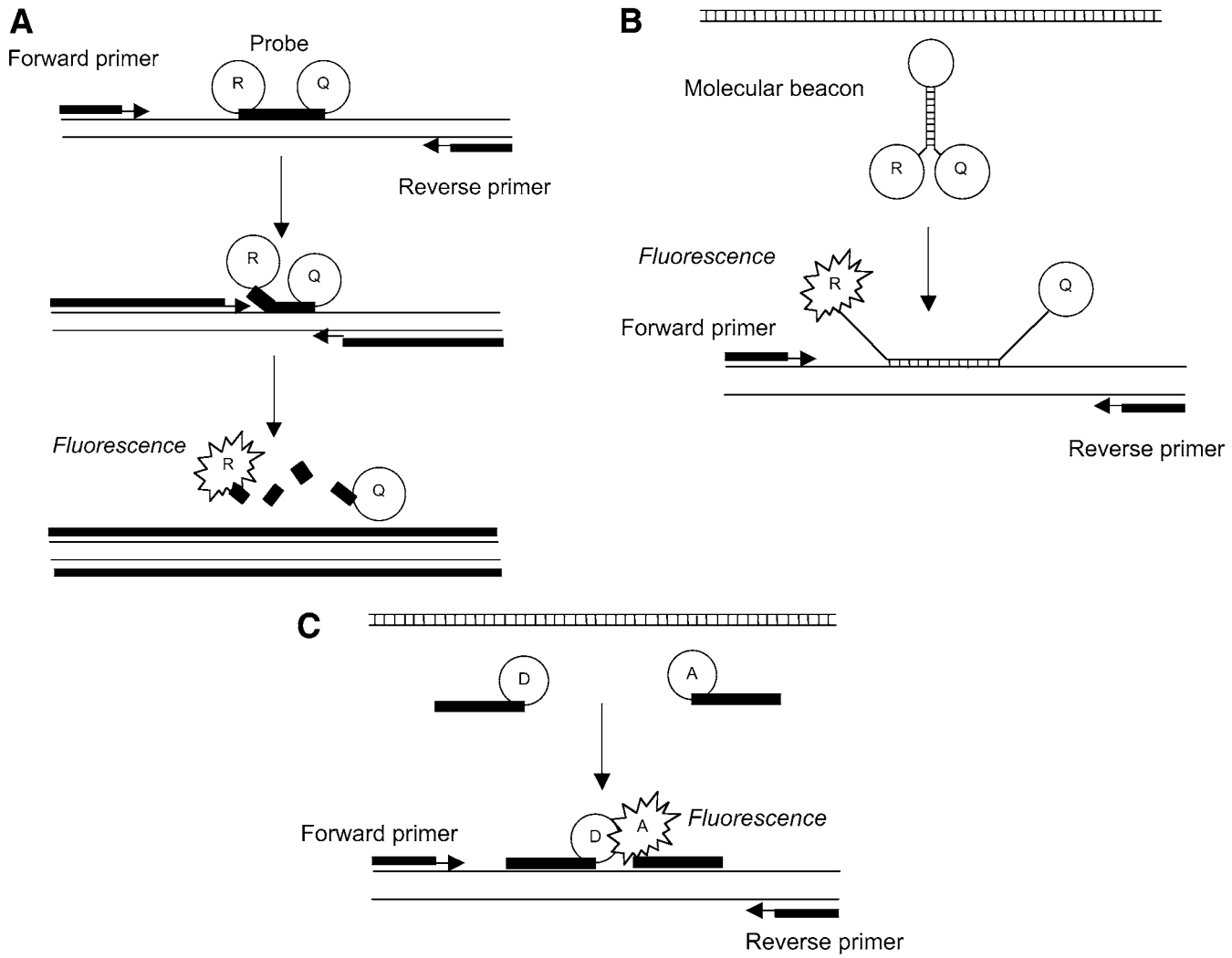


Fig. 1. Sequence-specific real-time PCR methods. **(A)** Cleavage (5′ exonuclease)-based assay: Light emission from the reporter fluorophore (R) is quenched because of its proximity to the quencher (Q). Cleavage by *Taq* polymerase separates the reporter and quencher allowing fluorescence. **(B)** Molecular beacons: Light emission from the reporter fluorophore (R) is quenched because of its proximity to the quencher (Q), brought about by the self-complementary 5′ and 3′ ends of the molecular beacon, causing it to take up a hairpin loop structure. Thermal denaturation and annealing allows the central loop section of the molecular beacon to bind to its target sequence in the PCR amplicon. Thus, the reporter and quencher become sufficiently separated to allow fluorescence. In this system, the probe is not subject to cleavage by *Taq* polymerase. **(C)** Fluorescence resonance energy transfer (FRET) probes: The acceptor flurophore (A) is unable to floresce until it is within 1–5 bp of the donor fluorophore (D), which occurs when the two FRET probes anneal to the specific PCR amplicon. **(D)** Scorpion primers: Light emission from the reporter fluorophore (R) of the Scorpion primer is quenched because of its proximity to the quencher (Q), brought about by the self-complementary of the stem sections, causing it to take up a hairpin loop structure. Thermal denaturation, annealing, and polymerase extension allows the central loop section of the Scorpion primer to bind to its target sequence in the PCR amplicon, thus stabilizing separation of the quencher and fluorescence. Thus, the reporter and quencher become sufficiently separated to allow fluorescence. In this system, the probe is not subject to cleavage by *Taq* polymerase. *Reverse primer and target DNA strand omitted for clarity. **(E)** Duplex Scorpion primers: Light emission from the reporter fluorophore (R) is quenched because of its proximity to the quencher (Q), brought about by the complementarity of the two Scorpion primers; the probe primer has the flourophore attached (R). Thermal denaturation separates the two primers allowing fluorescence, but annealing brings them back together again, causing quenching. However, in subsequent cycles, annealing of the probe sequence to its intramolecular complement in the amplicon is favored over binding to the quencher primer. Thus, the reporter and quencher became sufficiently separated to allow fluorescence. In this system, the probe is not subject to cleavage by *Taq* polymerase.

the 3′ end. When the probe is intact, the proximity of the quencher reduces the fluorescence emitted by the reporter dye. During PCR, the forward primer is extended by the *Taq* DNA polymerase until it reaches the probe. At this point, the exonuclease activity of the *Taq* DNA polymerase displaces and cuts up the probe, releasing the reporter dye and the quencher. Once they are no longer in close proximity, the reporter dye emits fluorescence of a particular wavelength that can be detected *(11)*.

2.2.2. Molecular Beacons Molecular beacons are self-complementary single-stranded oligonucleotides that form a hairpin loop structure (Fig. 1B) *(12,13)*. They consist of a probe homologous to the target sequence, flanked by sequences that are homologous to each other. Attached to one end is a reporter dye (FAM, TAMRA, TET, or ROX) and to the other is attached a quencher, usually DABCYL. When the beacon binds to the target sequence, the quencher and reporter are moved apart, and

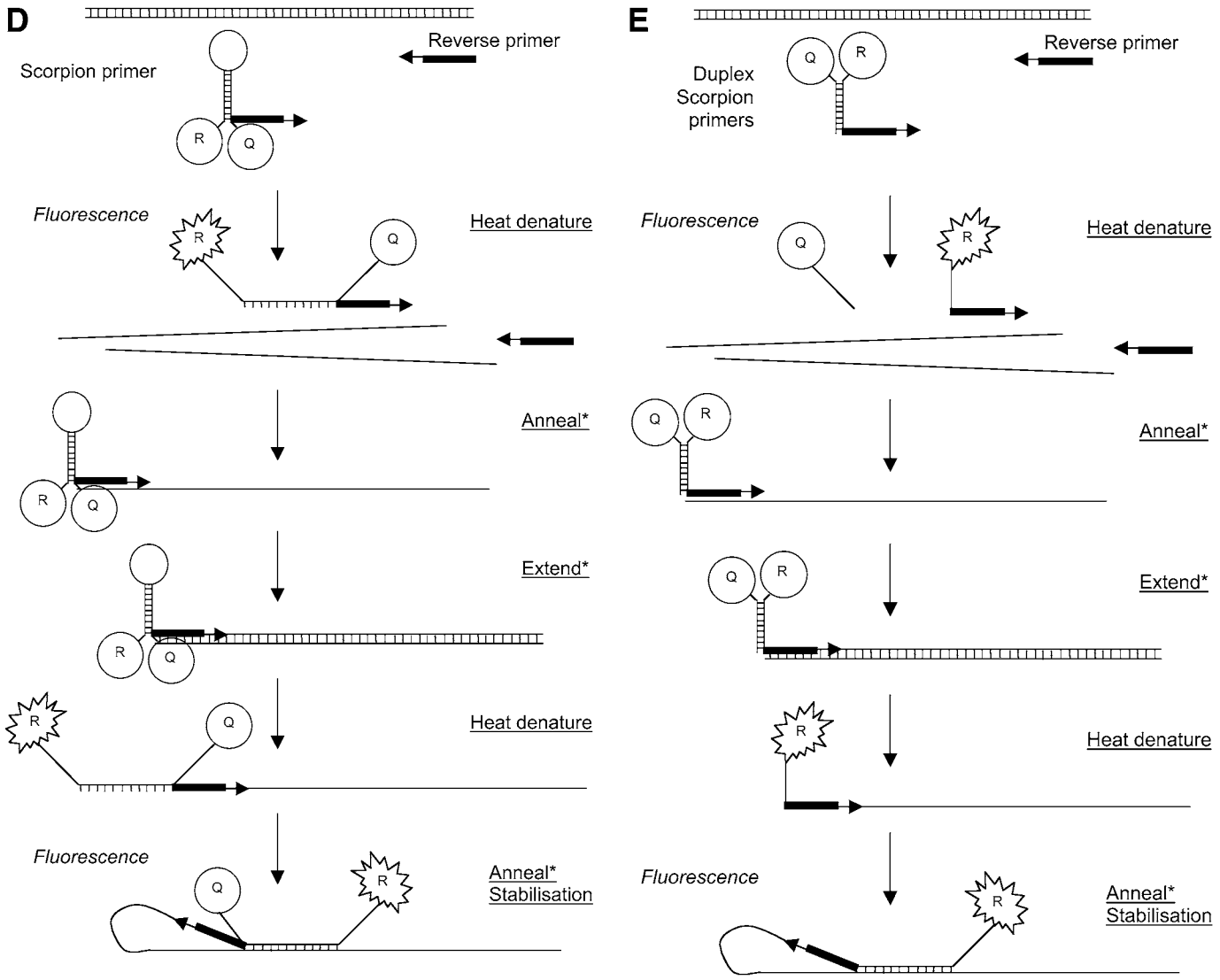


Fig. 1. (*Continued*).

fluorescence is emitted. Unlike in the cleavage-based assays, where fluorescence is detected during the elongation phase of PCR, with molecular beacons, fluorescence is detected during the annealing phase *(14–16)*.

2.2.3. FRET Probes Förster or fluorescence resonance energy transfer (FRET) probes are two separate fluorescently labeled oligonucleotides, one with a 5′ donor molecule and the other with a 3′ acceptor molecule attached: one is specific for target (i.e., wild-type or mutant) and the other is common (Fig. 1C). Only when these are placed within 1–5 bp of each other can energy be transferred from the donor to the acceptor, which then emits fluorescence *(17,18)*.

2.2.4. Scorpion Primers Scorpion primers are a form of molecular beacon, but unlike them, they act by a unimolecular rather than a bimolecular mechanism (Fig. 1D) *(19–21)*. Such a unimolecular mechanism is both faster and more efficient than a bimolecular mechanism, and direct comparison with equivalent cleavage (TaqMan) and molecular beacon-based tests would suggest that Scorpion primers perform best. By combining two Scorpion probes in multiplex PCR, it is possible to probe for two mutations simultaneously.

In its original form (Fig. 1D), a Scorpion primer has, from 5′ to 3′ the following: (1) a fluorophore at its 5′ end; (2) a stem and loop oligonucleotide sequence, the loop part of which is complementary to a sequence within the PCR amplicon; (3) a quencher dye (e.g., methyl red); (4) a PCR stopper/blocker section (e.g., hexethylene glycol, to prevent extension of the opposite strand beyond this point); (5) a conventional target-specific oligonucleotide PCR primer section *(22)*. After the initial PCR cycle, the loop section of the Scorpion probe is able to intramolecularly base-pair with its complementary sequence within the amplicon. The theoretical advantages of a Scorpion probe is that it is more specific, gives a stronger signal (improved signal-to-noise ratio), and acts faster than comparable molecular beacon or TaqMan probes. These theoretical advantages are borne out in practice.

Another version of the Scorpion probe system is the duplex format, whereby two primers complementary to each other, one

with a fluorophore and the other with a quencher, are used instead of a single stem-loop Scorpion probe (Fig. 1E) *(23)*. Like stem-loop Scorpions, this format requires the complex synthesis of nonstandard oligonucleotides *(22)*. However, the advantage of duplex Scorpions is that, like TaqMan probes, the quencher and fluorophore become completely separated during the process. This is unlike stem-loop Scorpions and molecular beacons, where some degree of Förster resonance energy transfer tends to quench the signal. The background is also theoretically lower compared with TaqMan probes because of higher quenching because of proximity of the fluorophore and quencher. This has led to the development of FRET duplex Scorpions *(23)*. Duplex Scorpion probes are easier to synthesize compared with stem-loop Scorpions, but, nonetheless, their synthesis is more involved than that for standard fluorescently labeled oligonucleotides.

So far, there have not been many medical applications of Scorpion probes have appeared, but those that have include analysis of known mutation detection in cystic fibrosis (*CFTR/ABCC7*), quantitation of human immunodeficiency virus (HIV)-1, and splice variant analysis of calpain 3 *(21,23–26)*. Some interesting applications to plant pathogen and food science have been presented *(27,28)*.

2.3. INFORMATION There is plenty of information on the Internet related to real-time PCR, in particular we would recommend the following: http://www.pcrlinks.com/variants/real-time_pcr.htm, http://www.protocol-online.org/prot/Molecular_Biology/PCR/Real-Time_PCR/, http://home. att.net/~dorak/genetics/realtime.html, and http://molecular-beacons.org/.

2.4. EQUIPMENT FOR REAL-TIME PCR Real-time-PCR is performed on thermal cyclers that have an integral optical system to detect fluorescence of different wavelengths. The optical system is connected to a computer with software that can analyze the fluorescence data to produce the exponential curves or allelic discrimination data required.

2.5. USES OF REAL-TIME PCR The main uses for real-time PCR in a molecular diagnostic laboratory are for single nucleotide polymorphism (SNP) genotype analysis and sequence copy number determination.

For SNP genotyping and small mutation testing, two differently labeled probes are designed: one for the wild-type allele and one for the mutant allele. The mismatch between the wild-type allele and the mutant probe facilitate competitive hybridization. Therefore, fluorescence will only be detected when the correct probe binds the target sequence. This sort of assay can be utilized where specific mutations are being analyzed, such as in sickle cell disease or as part of a cystic fibrosis screen *(11,14,17,18,20,23,29,30)*.

Real-time PCR can also be used to determine the copy number of specific target sequences, such as the Peripheral Myelin Protein 22 gene (*PMP22*). By multiplexing the primers and probes for *PMP22* with the primers and probes for a control sequence, known to be present in two copies, accurate measurement of the *PMP22* copy number can be made. This determines whether a patient has one, two, or three copies of *PMP22*. Patients with only one copy of *PMP22* have hereditary neuropathy with liability to pressure palsies (HNPP), whereas patients with three copies have Charcot–Marie–Tooth disease (CMT) *(3,24–26,31–33)*.

An interesting application of real-time PCR has been the detection of DNA methylation. Expression of the DNA mismatch repair enzyme *MLH1* is lost in approx 20% of colon cancers and is associated with methylation of the gene's promoter. Other tests for methylation, dependent on cleavage at or binding of primers to single specific CpG sites do not necessarily give results representative of the global methylation status of single DNA molecules. However, a real-time PCR assay (MethyLight), in which all the primers involved (forward, reverse, and reporter) bind to sites containing at least three CpGs each, has been designed *(34–36)*. Thus, the MethyLight assay only reports methylation when all of the CpG sites on a particular DNA molecule are methylated, which is likely to be more biologically relevant.

Real-time PCR also has many other applications in the fields of clinical microbiology, food microbiology, tumor biology, gene therapy, and gene expression *(4,27,28,37,38)*.

2.6. ADVANTAGES OF REAL-TIME PCR The main advantage of real-time PCR is the speed with which samples can be analyzed, as there are no post-PCR processing steps required. PCR can be performed in a 96- or 384-well format, and reactions can be multiplexed, leading to a high throughput of samples. The analysis of results is very simple and this contributes to it being a much faster and simpler method for analyzing gene copy number compared with many other current methods, such as multiplex amplifiable probe hybridization (MAPH), Southern blotting, and semiquantitative PCR *(39)*.

Another major advantage of having no post-PCR steps is that real-time PCR is a closed-tube method of analysis, which greatly reduces the chance of sample contamination, errors from mistakes in tube tranfers, or amplicons escaping into the laboratory environment; all important factors in the molecular diagnostics laboratory.

2.7. LIMITATIONS OF REAL-TIME PCR A major limitation of real-time PCR is the initial capital investment for equipment, as well as the investment required for staff training and expertise. Furthermore, real-time PCR probes are extremely expensive compared with conventional oligonucleotides. This underscores the fact that some of the personnel investment should be directed toward proper probe design. There are also technical limitations. When using nonspecific real-time PCR methods, such as with SYBR Green I, side-reaction products other than the desired amplicon (such as primer dimmers) are also detected. Careful optimization using melting-curve analysis is required to give a clear distinction between the correct amplicon and any unwanted PCR products. The inability to differentiate between specific and nonspecific products limits nonspecific real-time PCR to single amplicons; that is, the PCR cannot be multiplexed. Because there is a limited availability of fluorescent dye combinations, currently a maximum of four dyes can be used per reaction tube. This limits the extent to which multiplexing can be performed. A final limitation to be considered is that the quality and accuracy of the data produced are very dependent on the sample preparation and the quality of the DNA. Careful probe design and optimization are essential for good results to be obtained, particularly in SNP-based analyses. This means that careful planning and workup is required in order to set up a reliable and accurate assay. The inclusion of appropriate controls is mandatory.

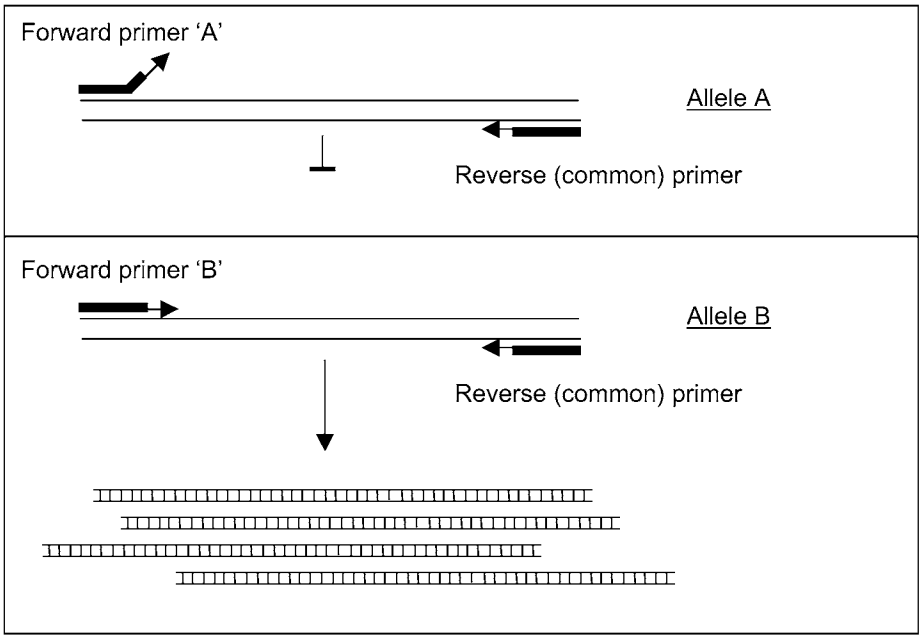


Fig. 2. The amplification refractory mutation system. The forward primer is designed at its 3′ end to match allele B, but not allele A. Thus, if allele B is present in the target DNA an amplicon will be produced in the PCR. A complementary ARMSPCR in which the forward primer is designed to match allele A, but not allele B, is also performed.

3. AMPLIFICATION REFRACTORY MUTATION SYSTEM

All PCRs depend on the annealing of oligonucleotide primers to specific sites in target DNA, facilitating amplification of a unique sequence. For the polymerase to extend the primers, they must be perfectly annealed to the target sequence at their 3′ ends. The amplification refractory mutation system (ARMS) exploits this by utilizing a primer that is designed such that its 3′ end matches, for example, one of two alternatives at a mutant nucleotide *(40)*. ARMS is ideally suited to the detection of point mutations and small insertions/deletions, and at its simplest, it involves two PCRs, one for each allele (Fig. 2). Thus, an ARMS PCR is allele-specific and earlier alternative names for the process have reflected this, including Allele-specific PCR (ASPCR) and PCR amplification of specific alleles (PASA) *(41,42)*. A number of ARMS variants have been produced and these are discussed in the following subsections.

Additional mismatches near the 3′ end of ARMS primers helps to increase the specificity of reactions, but too many mismatches can cause excessive destabilization. Differentiation of alleles is said to be problematic when the 3′ mismatch is C : A or involves a T, but this is not our experience *(43)*. The most discriminatory mismatches appear to be A : G/G : A, closely followed by Py : Py. It is wise to test the performance of a new ARMS PCR by varying reaction components and parameters, including primer concentration, buffer composition, magnesium concentration, annealing temperature, and PCR additives such as dimethyl sulfoxide (DMSO). Hot-start *Taq* is worth the extra cost. As always, novel applications require rigorous working up to test specificity, sensitivity, and other parameters of interest. It is essential to include a control PCR, that is not allele-specific, in the same tube as the ARMS reaction to avoid a null result. If a number of ARMS reactions are multiplexed, then this requirement is obviated.

3.1. DOUBLE ARMS Double ARMS utilizes a pair of allele-specific primers, instead of one specific and one common primer, as included in the standard ARMS test. The two polymorphic or mutant sites must be close enough to each other to allow a PCR to work, although given the distance that can be spanned by a long-range PCR (several kilobases), this is not the constraint it perhaps once was. Double ARMS enables the haplotyping of an individual in the absence of DNA from relatives. To do this, four ARMS PCRs must be set up: If locus 1 has two alleles (A and B) and locus 2 has two alleles (C and D), then these PCRs will be A × C, A × D, B × C, and B × D. Double ARMS can be useful for haplotyping doubly heterozygous individuals if two mutations are in cis or trans and/or to distinguish a carrier of a recessive disorder from an individual at risk of being affected. It has been used to haplotype alleles at closely linked loci, such as human leukocyte antigen (HLA) and rhesus blood groups *(44)*.

3.2. MS-PCR In some diagnostic applications, it is possible to include both allele-specific primers, one for each alternative at the mutant or polymorphic site, which then together with the common primer allows each amplicon to act as a control for the other. In this instance, there must be a way of distinguishing the two amplicons, either by differential size or labeling *(45–49)*. Advantages of MS-PCR include the fact that only a single PCR needs to be carried out and there is no need for a separate internal control. Sensitivity is improved because of the competitive nature of the PCR. Design of MS-PCR primers is somewhat more involved than simple ARMS assays, but this is not a significant negative factor.

3.3. MULTIPLEX ARMS In conditions where there are a restricted number of common mutations, then it might be efficacious to perform a test that simultaneously tests for as many of

them as possible (multiplex ARMS). This is exemplified by commercially available kits to detect particular *CFTR* mutations that are prevalent in particular populations. For instance, the Elucigene CF20 kit detects 20 different mutations in the *CFTR* gene *(44,49,50)*. It can be a difficult technical challenge to design primers that do not interfere with one another but still work in a multiplex ARMS. Amplicons are best distinguished by being either of differing sizes, or differentially labeled, or both.

3.4. ADVANTAGES AND LIMITATIONS OF ARMS The ARMS assays are generally quick, inexpensive, and simple to devise and work up. Inherently, they are not suited to screening for unknown mutations, but only mutations detected initially by some other technique. They do not require any unconventional PCR equipment, although the method used to detect amplicons can vary acording to that locally available or desired (from simple agarose gel electrophoresis to detection on an automated capillary fluorescent DNA analyzer). ARMS reactions are reasonably easily multiplexed and can also be used to determine haplotypes *(51)*. In terms of sensitivity, they are generally able to detect 1 mutant allele in 40 normal alleles. As with all oligonucleotide-based tests, care must be taken at the design stage to avoid known polymorphisms that might affect primer binding (such as *CFTR* ΔF508); similarly, caution must be exercised in interpretation, bearing in mind that a hitherto unknown sequence variant could exist in a patient and it interferes with or appears to be a known mutation *(52)*.

4. QUANTITATIVE FLUORESCENT PCR

Quantitative fluorescent PCR (QF-PCR) is based on the principle that the amount of amplicon produced by a PCR in its exponential phase (before the yield has reached its plateau) is proportional to the starting amount of target *(53)*. Thus, such PCRs are carried out with generally no more than 24 cycles. Fluorescent detection of such amplicons, via labeling of one of the primers, facilitates detection and quantitation of the PCR product using a fluorescent DNA analyzer, either gel or capillary based. No processing of the PCRs is necessary, other than the usual preparation of samples for running on such analyzers. Validation of such assays using samples tested by an independent method is necessary.

4.1. USES OF QF-PCR Quantitative fluorescent PCR is well suited to the rapid detection of aneuploidies in antenatal amniocentesis or chorion villus samples and is in wide use *(54–57)*. Fluorescence *in situ* hybridization (FISH) is perhaps slightly less reliable than QF-PCR for the detection of aneuploidy in such situations, and QF-PCR also has the benefit of easy automation. However, although FISH reagents are considerably more expensive than those needed for QF-PCR, the capital cost of a fluorescent microscope is less than a fluorescent DNA analyzer, although neither can be described as inexpensive *(58)*. QF-PCR has been applied to RhD testing, Duchenne muscular dystrophy testing, and analysis of *TP53* deletions in Li–Fraumeni syndrome *(59–61)*.

5. OLIGONUCLEOTIDE LIGATION ASSAY

Oligonucleotide primers that which have annealed to adjacent sites on single-stranded DNA, such that there is no gap between them, can be covalently joined together by DNA ligase.

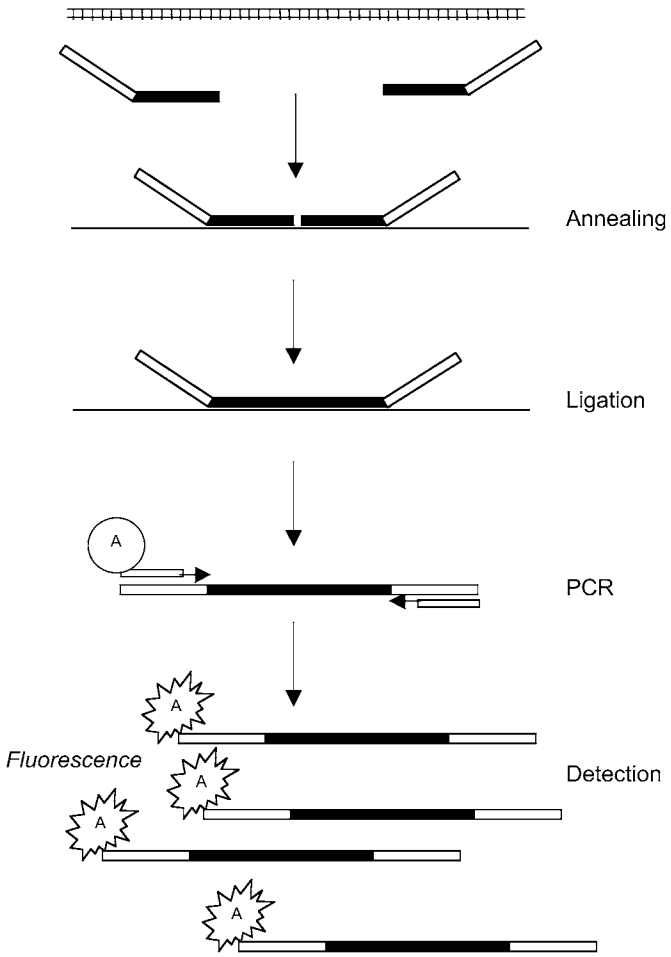


Fig. 3. Multiplex ligation-dependent probe amplification.

This is the basis for the oligoligation assay (oligonucleotide ligation [OLA]). Successive cycles of denaturation, annealing, and ligation result in a linear amplification of the target sequence. In contrast, standard PCR amplification produces an exponential increase in the target sequence. The adjacent ends of the primers must be perfectly matched with the target sequence, otherwise ligation will not occur. Thus, a single nucleotide or a few adjacent nucleotides can be interrogated *(62,63)*. Multiplexing allows a number of nucleotides and/or mutations to be detected in a single-reaction tube *(64)*. Currently, one of the primers usually has a fluorescent label so that the product can conveniently be detected on a fluorescent DNA analyzer. Although OLA is not strictly a PCR-based method, a new method has been developed in which an initial OLA is followed by a PCR, called multiplex ligation-dependent probe amplification (MLPA). As described next, MLPA is ideally suited to dosage testing of multiexon genes. A similar technique, in which the initial OLA is replaced by a PCR (universal primer quantitative fluorescent multiplex [UPQFM]) has also been described *(65)*.

5.1. MULTIPLEX LIGATION-DEPENDENT PROBE AMPLIFICATION Multiplex ligation-dependent probe amplification is a hybrid technique that combines an OLA with a QF-PCR. It is ideally suited to the measurement of exon copy number in multiexon genes, but it could be used for the quantitative assessment of any locus *(66)*.

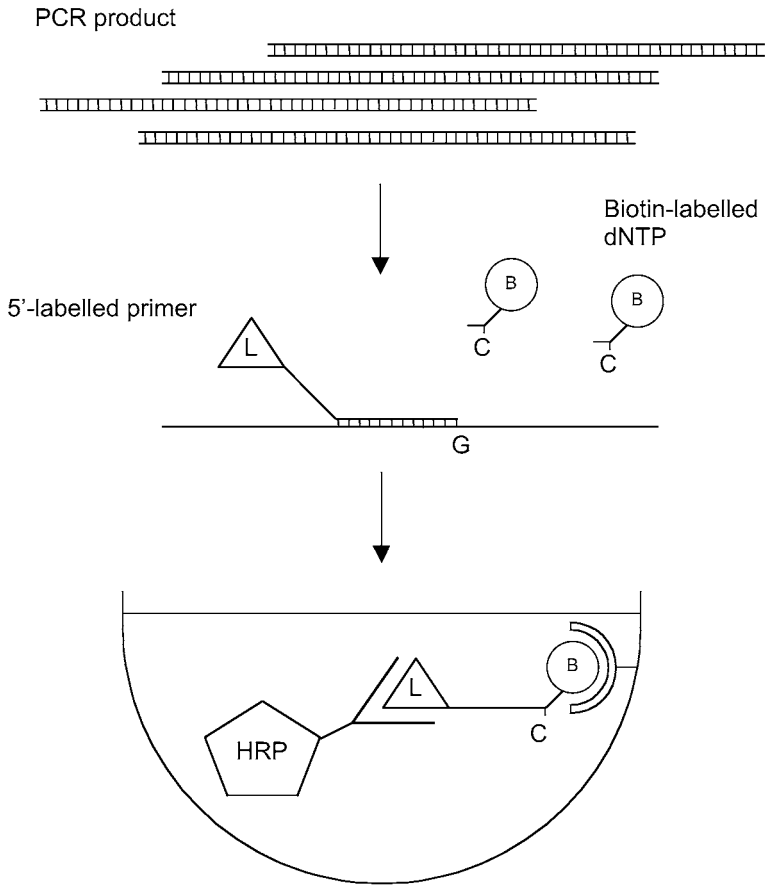


Fig. 4. Pronto assay. A conventional PCR is carried out, followed by enzymatic degradation of unincorporated dNTPs. A 5′-labeled primer is then annealed to the PCR and DNA polymerase and a biotinylated dNTP added. The biotinylated dNTP is only added to the primer if it matches the target amplicon. The reaction mix is then added to a streptavidin-coated well and biotinylated primers are thus bound; unbound primers are washed away. A HRP-linked antibody to the 5′ primer label is then added and detected by color development measured visually or by an ELISA plate reader.

In MLPA, genomic DNA is denatured and a mixture of oligonucleotide probes is hybridized to the DNA (Fig. 3). Each MLPA probe consists of two oligonucleotides that take part in a one-cycle OLA reaction. The two oligonucleotides have generic sequences attached to the 5′ end of the upstream primer and to the 3′ end of the downstream primer. Thus, after the initial OLA step, a PCR is then carried out using primers complementary to the generic sequences on the ends of the OLA primers. In this way, multiple different target sequences can be interrogated, but which are all amplified using the same PCR primers, with equal efficiency. As one of the generic primers is fluorescently labeled, then the fluorescence resulting from a particular amplicon (MLPA probe) is proportional to the amount of starting target DNA, and the copy number can be determined. The lengths of the various probes are engineered so that individual probes differ by a few basepairs and can be easily separated on a fluorescent DNA analyzer *(66)*.

5.2. ADVANTAGES AND USES OF MLPA Multiplex ligation-dependent probe amplification is quick, inexpensive and relatively simple. Probes can be easily designed for novel applications, and a wide range of clinically useful kits are commercially available (*see* http://www.mrc-holland.com/). The use of MLPA in detecting exon copy number changes in DNA mismatch repair genes in hereditary non-polyposis colorectal cancer (HNPCC) has been described *(67–69)*.

5.3. LIMITATIONS OF MLPA Kits are currently manufactured by only a single commercial company, although they are open to suggestions for new applications. Like any dosage technique, MLPA is sensitive to DNA quality and quantity. It is best to perform tests in duplicate and be suspicious of discordant results. Control probes are included in MLPA kits from MRC-Holland and notice should be taken of them and the manufacturer's instructions.

If a point mutation or polymorphism occurs within an MLPA probe-binding site, it might result in failure of ligation and, hence, appear as a deletion of a whole exon. The deletion of two or more contiguous exons is unlikely to be the result of this effect, but apparent deletion of a single exon might be. For this reason, another complementary technique should be used to confirm the nature of the mutation if it appears to be deletion of a single exon. This could be by using an alternative probe (pair of MLPA primers), real-time or long-range PCR, Southern blot, or mRNA analysis. It might also be useful to sequence the exon in question.

6. PRIMER EXTENSION

The single-nucleotide primer extension assay is in widespread use. It is the basis of a proprietary technique: Pronto *(70)* (Fig. 4). An amplicon is produced in a conventional PCR and residual dNTPs are eliminated using alkaline phosphatase.

The amplicon is denatured and a specific primer binds to one of the two DNA strands. At its 5′ end, the primer has a covalently linked label. Depending on the target sequence, a biotinylated dNTP is added to the 3′ end of the primer by DNA polymerase. The now biotinylated primer can bind to streptavidin-coated wells in a 96-well plate, and after washing, a horseradish peroxidase (HRP)-conjugated antibody to the primer label is added. Binding of the HRP-conjugated antibody is manifest by color development in that well and the mutation causing addition of the specific dNTP is detected. Different fluorescent labels can be linked to different dNTPs, such that with an appropriate detector, different mutations at the same nucleotide can be distinguished (such as in K-*ras*).

6.1. EQUIPMENT FOR PRIMER EXTENSION Aside from a conventional PCR thermocycler, results can be scored either visually or with an enzyme-linked immunosorbent assay (ELISA) plate reader if a numerical output is desired.

6.2. USES OF PRIMER EXTENSION In common with other PCR-based methods of mutation detection, primer extension is suited to the detection of defined mutations. The commercially available tests using Pronto concentrate on common point mutations in both dominant and recessive disorders, in particular those prevalent in the Ashkenazi Jewish population *(70–72)*.

6.3. ADVANTAGES AND LIMITATIONS OF PRIMER EXTENSION Primer extension is a gel-free mutation-detection system of low to high throughput. It only requires commonly available equipment. Although visual scoring is probably adequate for low-throughput applications, medium- to high-throughput benefits from using an ELISA reader.

Limitations include that Pronto kits are currently only manufactured by a single commercial company (http://www.savyondiagnostics.com/), a restricted range of applications is available, mostly based on conditions because of founder mutations in the Ashkenazi Jewish and Mediterranean populations *(70–72)*. However, a large number of in-house primer extension assays have been developed *(73,74)*. Like most PCR-based methods of mutation detection, primer extension is only able to detect defined mutations found by another technique; that is, it is not a method able to screen amplicons for any mutation.

REFERENCES

1. Aarskog, N. K. and Vedeler, C. A. Real-time quantitative polymerase chain reaction. *Hum. Genet.* **107:**494–498, 2000.
2. Klein, D. Quantification using real-time PCR technology: applications and limitations. *Trends Mol. Med.* **8:**257–260, 2002.
3. Zimmermann, B., Holzgreve, W., Wenzel, F., and Hahn, S. Novel real-time quantitative PCR test for trisomy 21. *Clin. Chem.* **48:**362–363, 2002.
4. Yang, J. H., Lai, J. P., Douglas, S. D., Metzger, D., Zhu, X. H., and Ho, W. Z. Real-time RT-PCR for quantitation of hepatitis C virus RNA. *J. Virol. Methods* **102:**119–128, 2002.
5. Emig, M., Saussele, S., Wittor, H., et al. Accurate and rapid analysis of residual disease in patients with CML using specific fluorescent hybridization probes for real time quantitative RT-PCR. *Leukemia* **13:**1825–1832, 1999.
6. Wittwer, C. T., Herrmann, M. G., Gundry, C. N., and Elenitoba-Johnson, K. S. Real-time multiplex PCR assays. *Methods* **25:**430–442, 2001.
7. van der Velden, V. H., Hochhaus, A., Cazzaniga, G., et al Detection of minimal residual disease in hematologic malignancies by real-time quantitative PCR: principles, approaches, and laboratory aspects. *Leukemia* **17:**1013–1034, 2003.
8. Wilhelm, J., Reuter, H., Tews, B., Pingoud, A., and Hahn, M. Detection and quantification of insertion/deletion variations by allele-specific real-time PCR: application for genotyping and chimerism analysis. *Biol. Chem.* **383:**1423–1433, 2002.
9. Tiemann, C., Vogel, A., Dufaux, B., Zimmer, M., Krone, J. R., and Hagedorn, H. J. Rapid DNA typing of HLA-B27 allele by real-time PCR using lightcycler technology. *Clin. Lab*. **47:**131–134, 2001.
10. Donohoe, G. G., Laaksonen, M., Pulkki, K., Ronnemaa, T., and Kairisto, V. Rapid single-tube screening of the C282Y hemochromatosis mutation by real-time multiplex allele-specific PCR without fluorescent probes. *Clin. Chem.* **46:**1540–1547, 2000.
11. Livak, K. J. Allelic discrimination using fluorogenic probes and the 5′ nuclease assay. *Genet. Anal.: Biomol. Eng.* **14:**143–149, 1999.
12. Abravaya, K., Huff, J., Marshall, R., et al. Molecular beacons as diagnostic tools: technology and applications. *Clin. Chem. Lab. Med.* **41:**468–474, 2003.
13. Broude, N. E. Stem-loop oligonucleotides: a robust tool for molecular biology and biotechnology. *Trends Biotechnol.* **20:**249–256, 2002.
14. Tapp, I., Malmberg, L., Rennel, E., Wik, M., and Syvanen, A. C. Homogeneous scoring of single-nucleotide polymorphisms: comparison of the 5′-nuclease TaqMan assay and Molecular Beacon probes. *Biotechniques* **28:**732–738, 2000.
15. Tsourkas, A., Behlke, M. A., Rose, S. D., and Bao, G. Hybridization kinetics and thermodynamics of molecular beacons. *Nucleic Acids Res.* **31:**1319–1330, 2003.
16. Ramachandran, A., Zhang, M., Goad, D., Olah, G., Malayer, J. R., and El-Rassi, Z. Capillary electrophoresis and fluorescence studies on molecular beacon-based variable length oligonucleotide target discrimination. *Electrophoresis* **24:**70–77, 2003.
17. Neoh, S. H., Brisco, M. J., Firgaira, F. A., Trainor, K. J., Turner, D. R., and Morley, A. A. Rapid detection of the factor V Leiden (1691 G > A) and haemochromatosis (845 G > A) mutation by fluorescence resonance energy transfer (FRET) and real time PCR. *J. Clin. Pathol.* **52:**766–769, 1999.
18. Wabuyele, M. B., Farquar, H., Stryjewski, W., et al. Approaching real-time molecular diagnostics: single-pair fluorescence resonance energy transfer (spFRET) detection for the analysis of low abundant point mutations in K-ras oncogenes. *J. Am. Chem. Soc.* **125:**6937–6945, 2003.
19. Whitcombe, D., Theaker, J., Guy, S. P., Brown, T, and Little, S. Detection of PCR products using self-probing amplicons and fluorescence. *Nat. Biotechnol.* **17:**804–807, 1999.
20. Whitcombe, D., Kelly, S., Mann, J., Theaker, J., Jones, C., and Little, S. Scorpion primers—a novel method for use in single tube genotyping. *Am. J. Hum. Genet.* **65:**2333, 1999.
21. Thelwell, N., Millington, S., Solinas, A., Booth, J., and Brown, T. Mode of action and application of Scorpion primers to mutation detection. *Nucleic Acids Res.* **28:**3752–3761, 2000.
22. Brown, L. J., McKeen, C. M., Mellor, J. M., Nicol, J. T., and Brown, T. Synthesis of fluorophore and quencher monomers for use in Scorpion primers and nucleic acid structure probes. *Organ. Biomol. Chem.* **1:**2267–2275, 2003.
23. Solinas, A., Brown, L. J., McKeen, C., et al. Duplex Scorpion primers in SNP analysis and FRET applications. *Nucleic Acids Res.* **29:**E96, 2001.
24. Cunnick, G. H., Jiang, W. G., Gomez, K. F., and Mansel, R. E. Lymphangiogenesis quantification using quantitative PCR and breast cancer as a model. *Biochem. Biophys. Res. Commun.* **288:** 1043–1046, 2001.
25. Saha, B. K., Tian, B., and Bucy, R. P. Quantitation of HIV-1 by real-time PCR with a unique fluorogenic probe. *J. Virol. Methods* **93:**33–42, 2001.
26. Taveau, M., Stockholm, D., Spencer, M., and Richard, I. Quantification of splice variants using molecular beacon or scorpion primers. *Anal. Biochem.* **305:**227–235, 2002.
27. Bates, J. A. and Taylor, E. J. A. Scorpion ARMS primers for SNP real-time PCR detection and quantification of *Pyrenophora teres*. *Mol. Plant Pathol.* **2:**275–280, 2001.

28. Terry, C. F., Shanahan, D. J., Ballam, L. D., Harris, N., McDowell, D. G., and Parkes, H. C. Real-time detection of genetically modified soya using Lightcycler and ABI 7700 platforms with TaqMan, Scorpion, and SYBR Green I chemistries. *J. AOAC Int.* **85:**938–944, 2002.
29. Frei, K., Szuhai, K., Lucas, T., et al. Connexin 26 mutations in cases of sensorineural deafness in eastern Austria. *Eur. J. Hum. Genet*. **10:**427–432, 2002.
30. Marras, S. A., Kramer, F. R., and Tyagi, S. Multiplex detection of single-nucleotide variations using molecular beacons. *Genet. Anal.* **14:**151–156, 1999.
31. Jebbink, J., Bai, X., Rogers, B. B., Dawson, D. B., Scheuermann, R. H., and Domiati-Saad, R. Development of real-time PCR assays for the quantitative detection of Epstein–Barr virus and cytomegalovirus, comparison of TaqMan probes, and molecular beacons. *J. Mol. Diagn*. **5:**15–20, 2003.
32. Li, Y., Zimmermann, B., Zhong, X. Y., Gupta, A. K., Holzgreve, W., and Hahn, S. Determination of RHD zygosity using real-time quantitative PCR. *Swiss Med. Wkly.* **133:**442–445, 2003.
33. Wilke, K., Duman, B., and Horst, J. Diagnosis of haploidy and triploidy based on measurement of gene copy number by real-time PCR. *Hum. Mutat.* **16:**431–436, 2000.
34. Trinh, B. N., Long, T. I., and Laird, P. W. DNA Methylation analysis by MethyLight technology. *Methods* **25:**456–462, 2001.
35. Eads, C. A., Danenberg, K. D., Kawakami, K., et al. MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res.* **28:**E32, 2000.
36. Eads, C. A., Lord, R. V., Wickramasinghe, K., et al. Epigenetic patterns in the progression of esophageal adenocarcinoma. *Cancer Res.* **61:**3410–3418, 2001.
37. Elsayed, S., Chow, B. L., Hamilton, N. L., Gregson, D. B., Pitout, J. D., and Church, D. L. Development and validation of a molecular beacon probe-based real-time polymerase chain reaction assay for rapid detection of methicillin resistance in Staphylococcus aureus. *Arch. Pathol. Lab. Med.* **127:**845–849, 2003.
38. Szuhai, K., Sandhaus, E., Kolkman-Uljee, S. M., et al. A novel strategy for human papillomavirus detection and genotyping with SybrGreen and molecular beacon polymerase chain reaction. *Am. J. Pathol*. **159:**1651–1660, 2001.
39. Hollox, E. J., Atia, T., Cross, G., Parkin, T., and Armour, J. A. High throughput screening of human subtelomeric DNA for copy number changes using multiplex amplifiable probe hybridisation (MAPH). *J. Med. Genet.* **39:**790–795, 2002.
40. Newton, C. R., Graham, A., Heptinstall, L. E., et al. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.* **17:**2503–2516, 1989.
41. Wu, D. Y., Ugozzoli, L., Pal, B. K., and Wallace, R. B. Allele-specific enzymatic amplification of beta-globin genomic DNA for diagnosis of sickle cell anemia. *Proc. Natl. Acad. Sci. USA* **86:**2757–2760, 1989.
42. Sommer, S. S., Cassady, J. D., Sobell, J. L., and Bottema, C. D. A novel method for detecting point mutations or polymorphisms and its application to population screening for carriers of phenylketonuria. *Mayo Clin. Proc*. **64:**1361–1372, 1989.
43. Kwok, S., Kellogg, D. E., McKinney, N., et al. Effects of primer–template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. *Nucleic Acids Res.* **18:**999–1005, 1990.
44. Chiu, R. W., Murphy, M. F., Fidler, C., Zee, B. C., Wainscoat, J. S., and Lo, Y. M. Determination of RhD zygosity: comparison of a double amplification refractory mutation system approach and a multiplex real-time quantitative PCR approach. *Clin. Chem.* **47:**667–672, 2002.
45. Rust, S., Funke, H., and Assmann, G. Mutagenically separated PCR (MS-PCR): a highly specific one step procedure for easy mutation detection. *Nucleic Acids Res.* **21:**3623–3629, 1993.
46. Schluter, B., Erren, M., Schotte, H., Junker, R., Rust, S., and Assmann, G. The mutagenically separated polymerase chain reaction is a rapid and reliable method for genotyping of the tumour necrosis factor-alpha promoter polymorphism (–308 G/A). *Clin. Chim. Acta.* **320:**135–138, 2002.
47. Purandare, S. M., Cawthon, R., Nelson, L. M., et al. Genotyping of PCR-based polymorphisms and linkage-disequilibrium analysis at the NF1 locus. *Am. J. Hum. Genet.* **59:**159–166, 1996.
48. Myint, L., Ariyoshi, K., Yan, H., et al. Mutagenically separated PCR assay for rapid detection of M41L and K70R zidovudine resistance mutations in CRF01_AE (subtype E) human immunodeficiency virus type 1. *Antimicrob. Agents Chemother.* **46:**3861–3868, 2002.
49. Chang, J. G., Liu, H. J., Huang, J. M., Yang, T. Y., and Chang, C. P. Multiplex mutagenically separated PCR: diagnosis of beta-thalassemia and hemoglobin variants. *Biotechniques* **22:**520–527, 1997.
50. Ferrie, R. M., Schwarz, M. J., Robertson, N. H., et al. Development, multiplexing, and application of ARMS tests for common mutations in the CFTR gene. *Am. J. Hum. Genet.* **51:**251–262, 1992.
51. Cheadle, J., Myring, J., al-Jader, L., and Meredith, L. Mutation analysis of 184 cystic fibrosis families in Wales. *J. Med. Genet.* **29:**642–646, 1992.
52. Kobayashi, K., Knowles, M. R., Boucher, R. C., O'Brien, W. E., and Beaudet, A. L. Benign missense variations in the cystic fibrosis gene. *Am. J. Hum. Genet.* **47:**611–615, 1990.
53. Butler, J. M., McCord, B. R., Jung, J. M., Wilson, M. R., Budowle, B., and Allen, R. O. Quantitation of polymerase chain reaction products by capillary electrophoresis using laser fluorescence. *J. Chromatogr. B. Biomed. Appl*. **658:**271–280, 1994.
54. Pertl, B., Yau, S. C., Sherlock, J., Davies, A. F., Mathew, C. G., and Adinolfi, M. Rapid molecular method for prenatal detection of Down's syndrome. *Lancet* **343:**1197–1198, 1994.
55. Adinolfi, M., Pertl, B., and Sherlock, J. Rapid detection of aneuploidies by microsatellite and the quantitative fluorescent polymerase chain reaction. *Prenat. Diagn.* **17:**1299–1311, 1997.
56. Pertl, B., Kopp, S., Kroisel, P. M., Tului, L., Brambati, B., and Adinolfi, M. Rapid detection of chromosome aneuploidies by quantitative fluorescence PCR: first application on 247 chorionic villus samples. *J. Med. Genet.* **36:** 300–303, 1999.
57. Mann, K., Fox, S. P., Abbs, S. J., et al. Development and implementation of a new rapid aneuploidy diagnostic service within the UK National Health Service and implications for the future of prenatal diagnosis. *Lancet* **358:**1057–1061, 2001.
58. Hulten, M. A., Dhanjal, S., and Pertl, B. Rapid and simple prenatal diagnosis of common chromosome disorders: Advantages and disadvantages of the molecular methods FISH and QF-PCR. *Reproduction* **126:**279–297, 2003.
59. Pertl, B., Pieber, D., Panzitt, T., et al. RhD genotyping by quantitative fluorescent polymerase chain reaction: a new approach. *BJOG* **107:**1498–1502, 2000.
60. Mansfield, E. S., Robertson, J. M., Lebo, R. V., et al. Duchenne/Becker muscular dystrophy carrier detection using quantitative PCR and fluorescence-based strategies. *Am. J. Med. Genet.* **48:**200–208, 1993.
61. Bougeard, G., Brugieres, L., Chompret, A., et al. Screening for TP53 rearrangements in families with the Li-Fraumeni syndrome reveals a complete deletion of the TP53 gene. *Oncogene* **22:**840–846, 2003.
62. Delahunty, C. M., Ankener, W., Brainerd, S., Nickerson, D. A., and Mononen, I. T. Finnish-type aspartylglucosaminuria detected by oligonucleotide ligation assay. *Clin. Chem.* **41:**59–61, 1995.
63. Brinson, E. C., Adriano, T., Bloch, W., et al. Introduction to PCR/OLA/SCS, a multiplex DNA test, and its application to cystic fibrosis. *Genet. Test.* **1:**61–68, 1997.
64. Grossman, P. D., Bloch, W., Brinson, E., et al. High-density multiplex detection of nucleic acid sequences: oligonucleotide ligation assay and sequence-coded separation. *Nucleic Acids Res.* **22:**4527–4534, 1994.
65. Heath, K. E., Day, I. N., and Humphries, S. E. Universal primer quantitative fluorescent multiplex (UPQFM) PCR: a method to detect major and minor rearrangements of the low density lipoprotein receptor gene. *J. Med. Genet.* **37:**272–280, 2000.
66. Schouten, J. P., McElgunn, C. J., Waaijer, R., Zwijnenburg, D., Diepvens, F., and Pals, G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* **30:**e57, 2002.

67. Hogervorst, F. B., Nederlof, P. M., Gille, J. J., et al. Large genomic deletions and duplications in the BRCA1 gene identified by a novel quantitative method. *Cancer Res.* **63:**1449–1453, 2003.
68. Montagna, M., Palma, M. D., Menin, C., et al. Genomic rearrangements account for more than one-third of the BRCA1 mutations in northern Italian breast/ovarian cancer families. *Hum. Mol. Genet.* **12:**1055–1061, 2003.
69. Nakagawa, H., Hampel, H., and de la Chapelle, A. Identification and characterization of genomic rearrangements of MSH2 and MLH1 in Lynch syndrome (HNPCC) by novel techniques. *Hum. Mutat*. **22:**258, 2003.
70. Horowitz, M., Pasmanik-Chor, M., Borochowitz, Z., et al. Prevalence of glucocerebrosidase mutations in the Israeli Ashkenazi Jewish population. *Hum. Mutat.* **12:**240–244, 1998. Erratum: *Hum. Mutat.* **13:**255, 1999.
71. Karpati, M., Peleg, L., Gazit, E., Akstein, E., and Goldman, B. A novel mutation in the HEXA gene specific to Tay–Sachs disease carriers of Jewish Iraqi origin *Clin. Genet.* **57:**398–400, 2000.
72. Drucker, L., Stackievitz, R., Shpitz, B., and Yarkoni, S. Incidence of BRCA1 and BRCA2 mutations in Ashkenazi colorectal cancer patients: preliminary study. *Anticancer Res.* **20:**559–561, 2000.
73. Jordan, N., Williams, N., Gregory, J. W., Evans, C., Owen, M., and Ludgate, M. The W546X mutation of the thyrotropin receptor gene: potential major contributor to thyroid dysfunction in a Caucasian population. *J. Clin. Endocrinol. Metab.* **88:**1002–1005, 2003.
74. Matyas, G., Giunta, C., Steinmann, B., Hossle, J. P., and Hellwig, R. Quantification of single nucleotide polymorphisms: a novel method that combines primer extension assay and capillary electrophoresis. *Hum. Mutat*. **19:**58–68, 2002.

8 Alternative Methods for Amplified Nucleic Acid Testing

DEBORAH A. PAYNE AND LAURIE E. SOWER

1. INTRODUCTION

Molecular pathology is a relatively new division of laboratory medicine that detects, characterizes, and/or quantifies nucleic acids to assist in the diagnosis of human disease. Molecular assays augment classical areas of laboratory medicine by providing additional diagnostic data either in a more expeditious manner or by providing results that would not be obtainable using standard methodologies. For these reasons, molecular pathology is the most rapidly growing area in laboratory medicine. Molecular pathology can be categorized into five subdivisions that specialize in the diagnosis of diseases or conditions associated with (1) hematology/oncology, (2) solid tumors, (3) genetics, (4) pharmacogenetics, and (5) infectious diseases. Based on test volume, detection, and characterization of infectious diseases is currently the dominant subdivision of molecular pathology and is projected to continue to dominate the other areas of molecular pathology for the next several years.

Molecular methods used for detecting infectious agents have several advantages when compared to classical microbiology approaches. Molecular methods are highly sensitive and, therefore, can detect minute amounts of infectious agents. Because these methods generally do not require growth in culture media, various bacteria, viruses, and fungi, which are difficult or impossible to culture, can be readily identified *(1–8)*. In addition, the sensitivity of these methods can allow the analysis of nonviable infectious agents (i.e., permitting the analysis of archived formalin-fixed tissue). The viral load and genotype of certain infectious agents that can facilitate treatment protocols can also be determined. These advantages are some of the reasons why infectious diseases are dominating the area of Clinical Molecular Pathology. As a result, these new molecular methodologies have placed the highest priority on the ability to equally detect and quantify genetic variants of various infectious agents.

Initially, polymerase chain reaction (PCR) was the primary molecular method used in clinical laboratories. However, numerous negative factors have limited its role in diagnostic testing laboratories. Some of these issues are PCR's requirement for (1) specialized instruments, (2) workflow, (3) specially trained personnel, (4) substantial quality assurance, and (5) overhead cost for intellectual properties. These issues, combined with the increase in the infectious diseases diagnostic market, have supported the development of numerous other molecular platforms. The success of these platforms in the infectious disease arena has resulted in a spillover into other areas of clinical molecular diagnostics, specifically genetic and solid tumor testing. Recent advances in genetic medicine is now requiring the development of methods that can detect and identify numerous genetic polymorphisms ranging from viral sequences to human single-nucleotide polymorphisms (SNPs) to whole genomic scanning *(9)*. Various platforms have been developed to address these goals and include technologies that range from single-temperature-based assays to multiple-temperature assays that can either amplify a target nucleic acid or a reporter molecule. The focus of this chapter will be to provide an overview of some of these non-PCR-based detection systems for nucleic acid testing (NAT).

2. TARGET VS SIGNAL AMPLIFICATION

Nucleic acids can be detected using either target or signal amplification methods. Briefly, target amplification enzymatically increases the number of target molecules. In short, "at the end of the day," there are more molecules of the targeted nucleic acid. In contrast, signal amplification does not increase the target but uses highly sensitive reporter molecules or probes to detect the target. For example, in signal amplification, the same number of molecules exists at the end of the day, but the molecular methods act as a kind of "magnifying glass" to aid in their "visualization."

2.1. TARGET AMPLIFICATION Target amplification is the most frequently used method and is accomplished using several technologies. Target amplification increases the amount of the infectious agent's nucleic acid in a test tube by employing an enzymatic in vitro replication step. Examples of target amplification techniques are PCR, transcription-mediated amplification (TMA), nucleic acid sequence based amplification (NASBA), rolling cycle amplification (RCA), ligase chain reaction (LCR), and strand displacement amplification (SDA). All of these

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

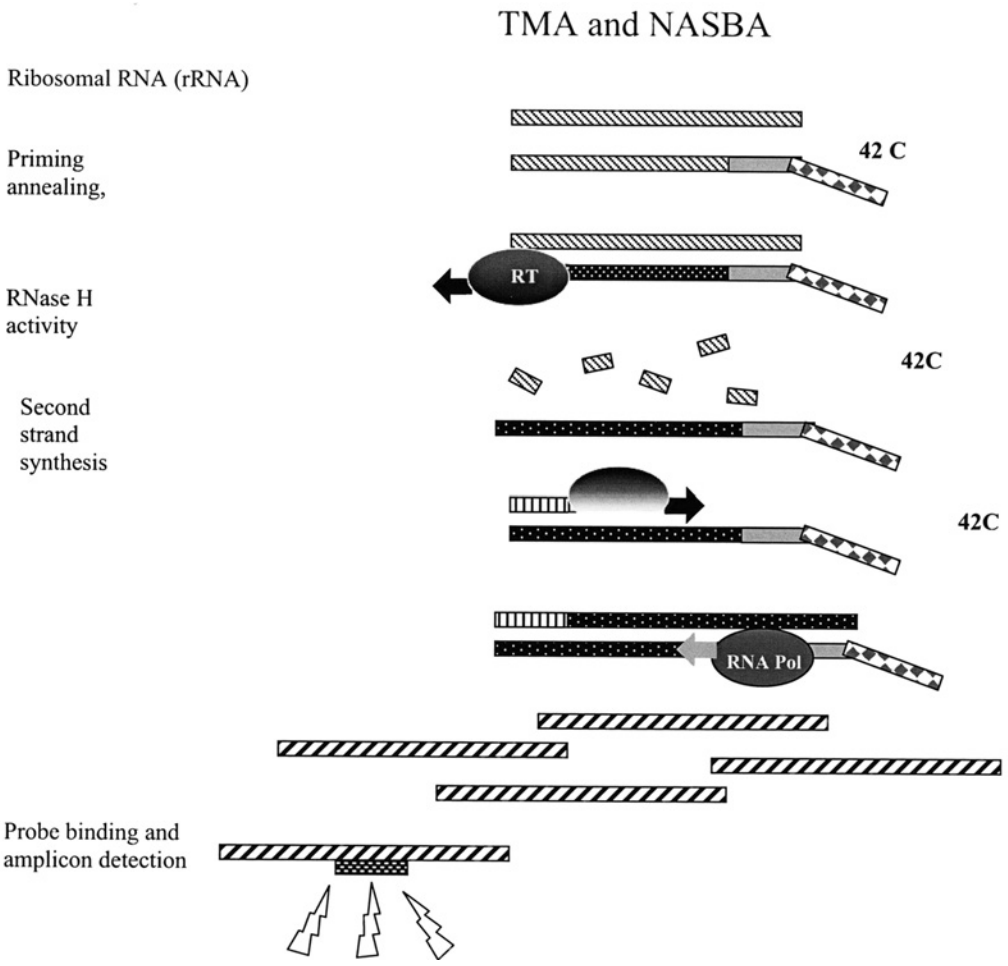


Fig. 1. Isothermal one dimensional target or probe.

methods use a polymerase or ligase and short synthetic oligonucleotides known as primers. The primers specifically bind to complementary sequences found in the infectious agent. In most of the commercially available kits, an additional level of specificity is added to the assay with a third probe that is complementary to the amplified target sequence. The probe is used to detect the amplicon. The main differences between these approaches are the use of multiple temperatures vs a single temperature and whether the target or probes are linear or circular.

2.2. ISOTHERMAL AMPLIFICATION Isothermal methods acquired their name after the Latin meaning of "iso" being same or similar and after the Greek "thermal (therme)" meaning heat. Isothermal amplification can be further subdivided into methods that target single-dimensional linear nucleic acid targets vs methods that require two-dimensional targets or probes (i.e., circular molecules). Isothermal approaches do not require thermocyclers and is one of the advantages of these methods. In addition, these methods have a high throughput because all steps can be performed in a single tube.

2.2.1. One-Dimensional or Linear Targets/Probes (TMA and NASBA) Figure 1 outlines TMA and NASBA amplification methods *(10,11)*. The most common nucleic acid targets for these techniques are ribosomal RNA (rRNA) sequences that exist in substantial numbers (e.g., exceeding 10,000 copies per organism). Targeting these nucleic acid sequences takes advantage of the naturally occurring multimeric copies of these sequences. To begin, a customized primer with a binding sequence for an RNA polymerase and a region that is complementary to the target rRNA is synthetically made. The region of this primer that is complementary to the rRNA sequence upon hybridization acts as a primer for generating a complementary DNA (cDNA) molecule using a reverse transcriptase enzyme. RNase H activity degrades the rRNA target, leaving the single strand of cDNA. The addition of a DNA polymerase and another complementary DNA primer generates the second strand. The final product results in a double-stranded DNA (dsDNA) copy of the original rRNA sequence plus an RNA polymerase-binding region. The addition of an RNA polymerase causes this dsDNA to act as a template for the generation of numerous RNA molecules via transcription (hence, the name transcription-mediated amplification). By targeting a high-copy rRNA molecule rather than a DNA sequence that is present at 1–2 copies per organism, every 10,000 rRNAs will be converted to a transcriptional active DNA template capable of producing numerous copies of RNA. Finally, a probe complementary to the amplicon is used to detect the amplified sequence. Although rRNA sequences are most often targeted, this approach can be used for other RNA targets and, with minor adjustments, for DNA targets as well.

2.2.2. Strand Displacement Amplification Like TMA and NASBA, SDA requires multiple enzymes (e.g., a thermostable polymerase and restriction enzyme) *(12–14)*.

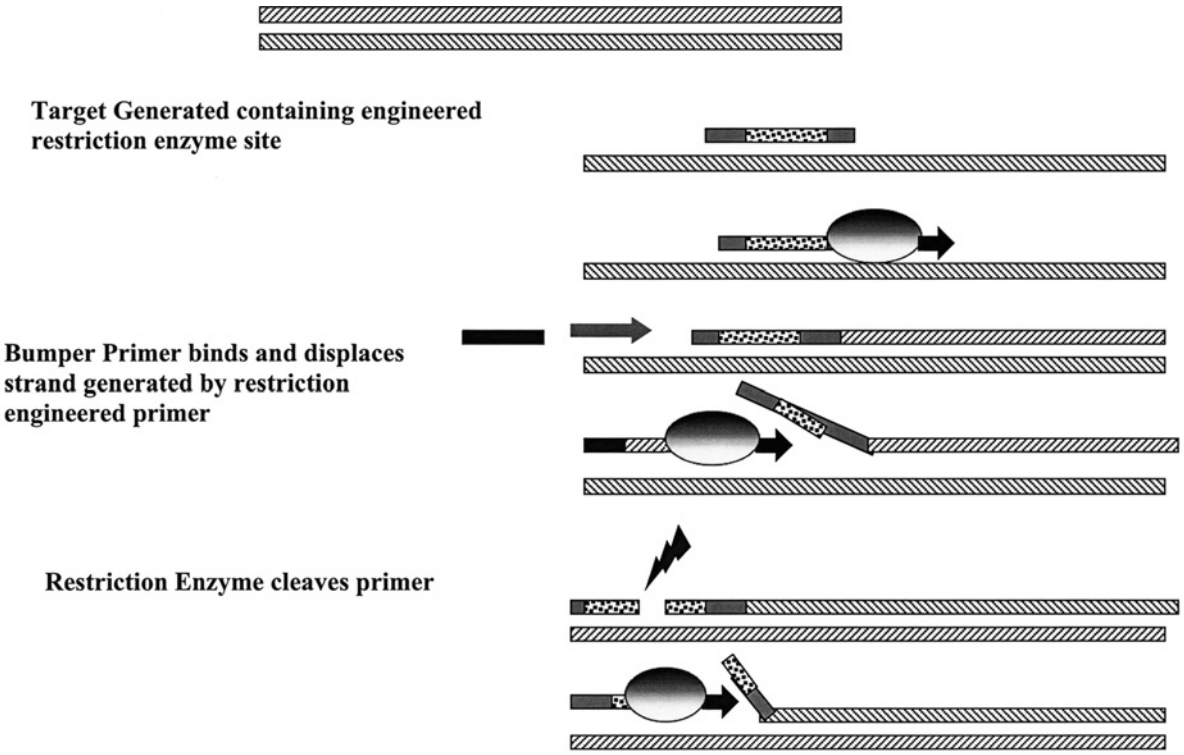


Fig. 2. Strand displacement amplification.

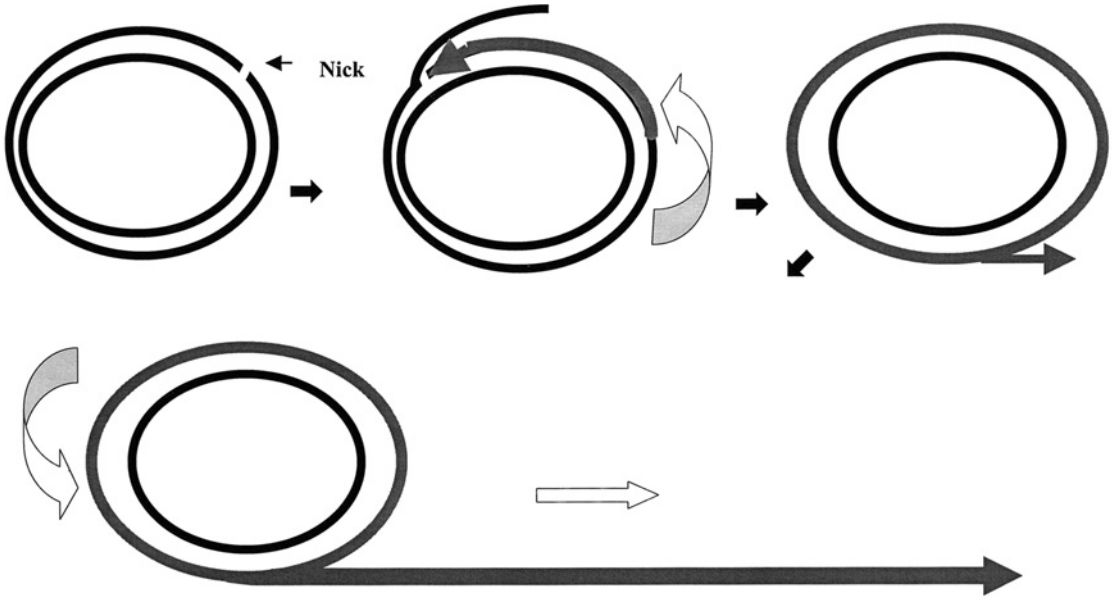


Fig. 3. Rolling circle amplification.

However, in contrast, it requires multiple primers in a specific order (four total) to amplify the target sequence and displace the copied sequence (Fig. 2). An additional difference is its use of a chemically modified deoxynucleotide base (thiolated dCTP). The amplification process uses two phases: the target generation phase and the amplification phase. In the target generation phase, an engineered primer that has a restriction enzyme site incorporated into it binds to its complementary target and initiates strand synthesis using a thermostable polymerase. A bumper primer displaces the strand generated from the primer containing the restriction enzyme site. Because the newly generated strands incorporate thiolated dCTP, they are not susceptible to restriction enzymatic digestion. A thermostable restriction enzyme introduces a single-strand nick in the double-stranded molecules. The thermostable polymerase then extends the new strand and thereby displaces the strand 3′ to the nick. Ultimately, new strands that incorporate this restriction enzyme site lead to the exponential generation of target copies.

2.2.3. Two-Dimensional or Circular Targets/Probes

The RCA mechanism of replication is depicted in Fig. 3. The replication is initiated by a single-stranded nick and begins

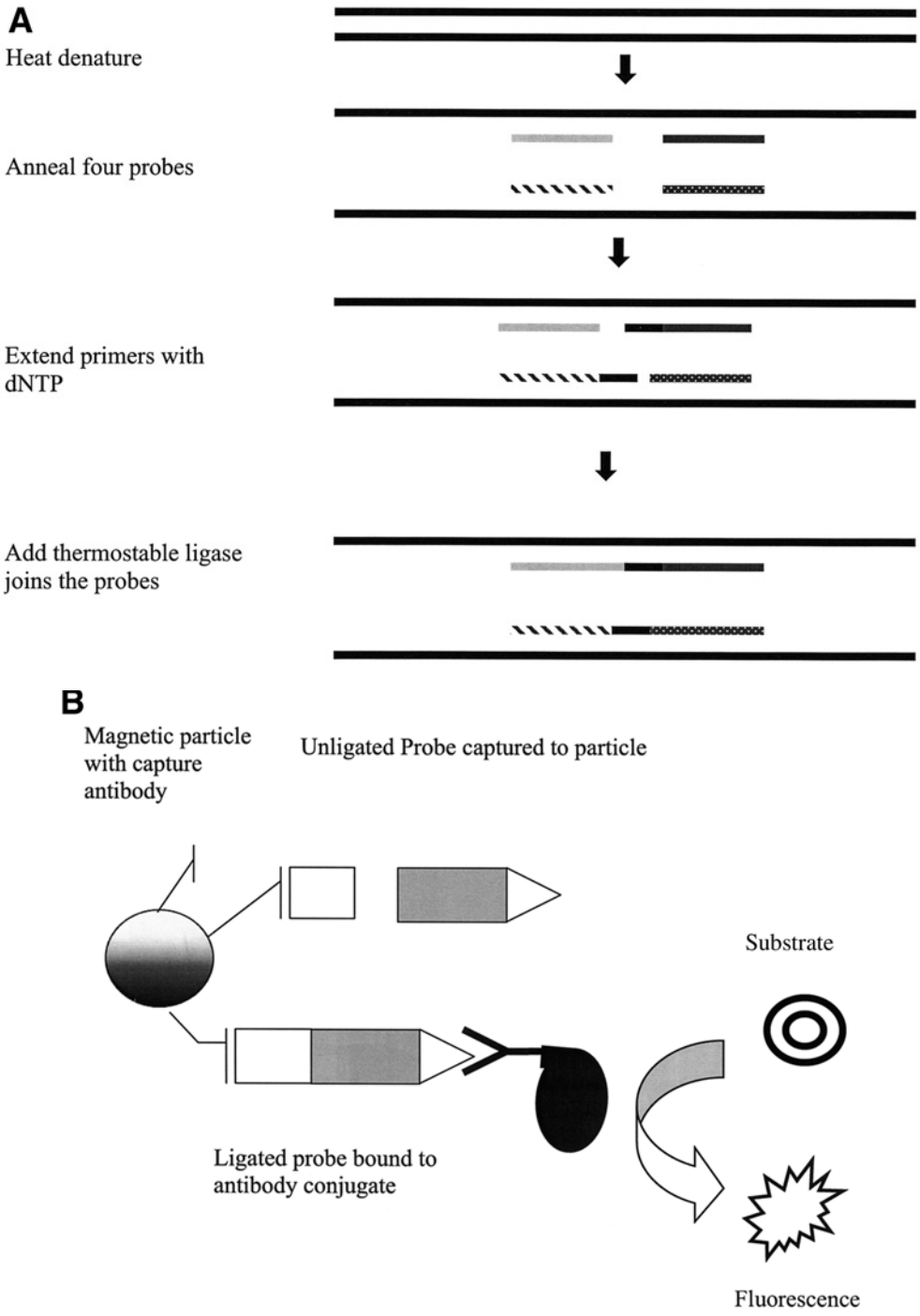


Fig. 4. (A) Ligase chain reaction; **(B)** ligase chain reaction detection.

with the extension of the strand upstream from the nick and proceeds 5′ to 3′.The portion of the nicked strand that is 3′ or downstream from the nick is displaced as the newly synthesized strand rolls around the circular DNA molecule. The result is a single-stranded tail that forms concatemeric units of the initial circle. Finally, a complementary strand is formed from the single-stranded unit. The rolling circle form of replication is used for the sensitive detection of single-molecule, single-base-point mutations and sequencing applications *(15–25)*. This method has also been useful for whole genomic amplification using high-fidelity and high-processitivity polymerases in the MDA *(9)*. This method uses multiple primers, which bind to the denatured DNA and can generate 33,000 ng of DNA from 0.3 ng of starting material. This method has been shown to replicate the entire genomic sequence in an unbiased fashion. As a result, MDA could prove to be extremely valuable in producing large amounts of rare control material for clinical laboratories.

2.3. MULTIPLE THERMAL AMPLIFICATION

2.3.1. Ligase Chain Reaction In multithermal target amplification (specifically LCR), a high temperature is used to denature or unzip the double-stranded DNA (dsDNA) target

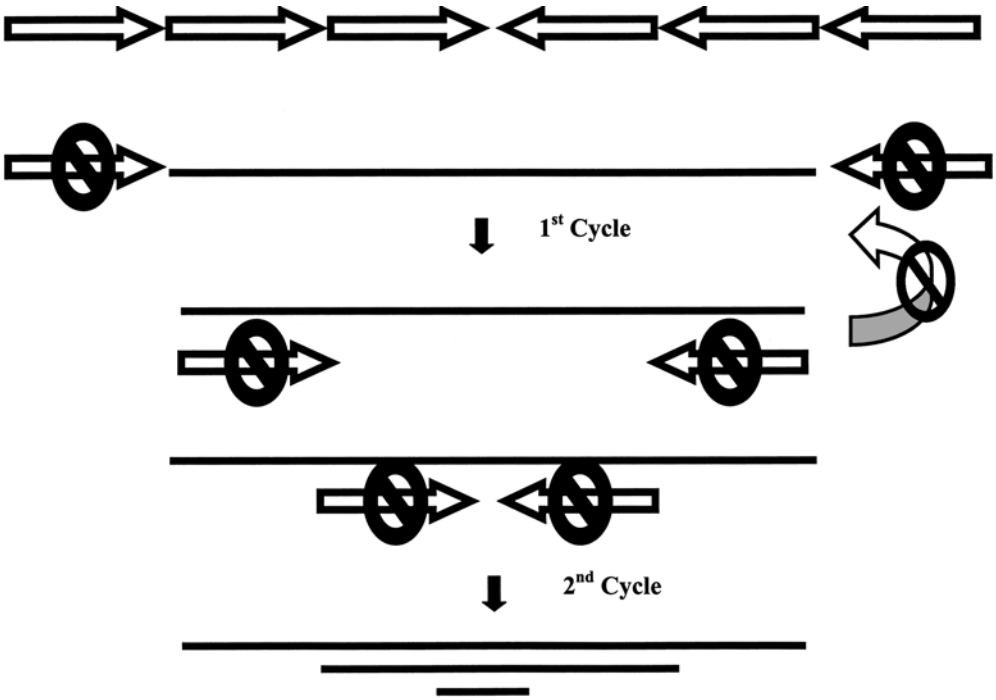


Fig. 5. linked linear amplification.

(26,27). Multiple probes complementary to different regions of the target then bind to their specific targets in the annealing step. The next step in the LCR reaction takes place at the optimal temperature for the DNA polymerase so that elongation or extension can proceed from the probes. Finally, a ligase joins the slightly extended probes together. Because LCR requires multiple cycles through all three temperatures, a thermostable enzyme capable of withstanding temperatures of 95°C and higher without becoming denatured or inactive is necessary. LCR results in the logarithmic amplification of a target sequence by doubling the target during each cycling step. The use of multiple probes in the reaction can negate potential problems arising from polymorphic regions that could adversely affect probe binding. The ligated product is then detected using a capture system. Specifically, a hapten is attached to the ends of the capture probe and the target-specific probe. A separate capture probe is removed from the reaction solution and all nonbond target probes are washed away. Only target probes bound to the capture probe can be detected using detection molecules directed to the hapten label on the 3′ end of the target molecule. Figure 4B demonstrates how these amplified products are captured and detected. The closed-tube system decreases the risk for contamination and the turn around time for the assay.

2.3.2. Linked Linear Amplification Linked linear amplification (Fig. 5) is a method that is very similar to PCR except for its use of multiple nested nonreplicable primers. The nonreplicable primers contain a molecule (1,3 propanediol) that prevents replication during the extension phase of the thermocycling reactions. Because the ends of the amplified strands cannot serve as templates for amplification, the products become progressively smaller. The main advantage of this method is its resistance to amplicon contamination because the progressively smaller products lack the primer sequences of the larger products. In addition, its ability to accommodate polymorphic targets through its use of multiple primers render it resistant to potential false-negative results resulting from polymorphic target sequences *(28)*.

3. SIGNAL AMPLIFICATION

Signal amplification methods have been available in serology for many years and are characterized by increases in the detection of a target molecule rather than the molecule itself. Similar principles are used when detecting various nucleic acid targets. Two approaches of signal amplification methods involve nonenzymatic or enzymatic technologies. In nonenzymatic signal amplification, the target is immobilized or localized using complementary probes. Ultimately, the hybrid molecule is detected using molecules labeled with alkaline phosphatase that can generate a colorimetric or chemiluminescence reaction (Fig. 6A and 6B). Alternatively, fluorescent probes can be used to directly visualize the hybrid molecule for *in situ* applications (Fig. 7). Signal amplification techniques are not prone to the problems of amplicon contamination because the target is not amplified. The other category of signal amplification uses enzymatic activities to replicate or generate a reporter molecule. These enzymatic signal amplification methods can be susceptible to false-positive reactions if the reporter molecule inadvertently contaminates the specimen. As a result, utmost care must be exercised when disposing of the specimen following detection.

3.1. NONENZYMATIC SIGNAL AMPLIFICATION

3.1.1. Hybrid Capture The hybrid capture kits from Digene Inc. (Gaithersburg, MD) can be used to detect HPV, CT/GC, and HSV. Full-length RNA probes are used for the detection of the HPVs and to determine whether they have a

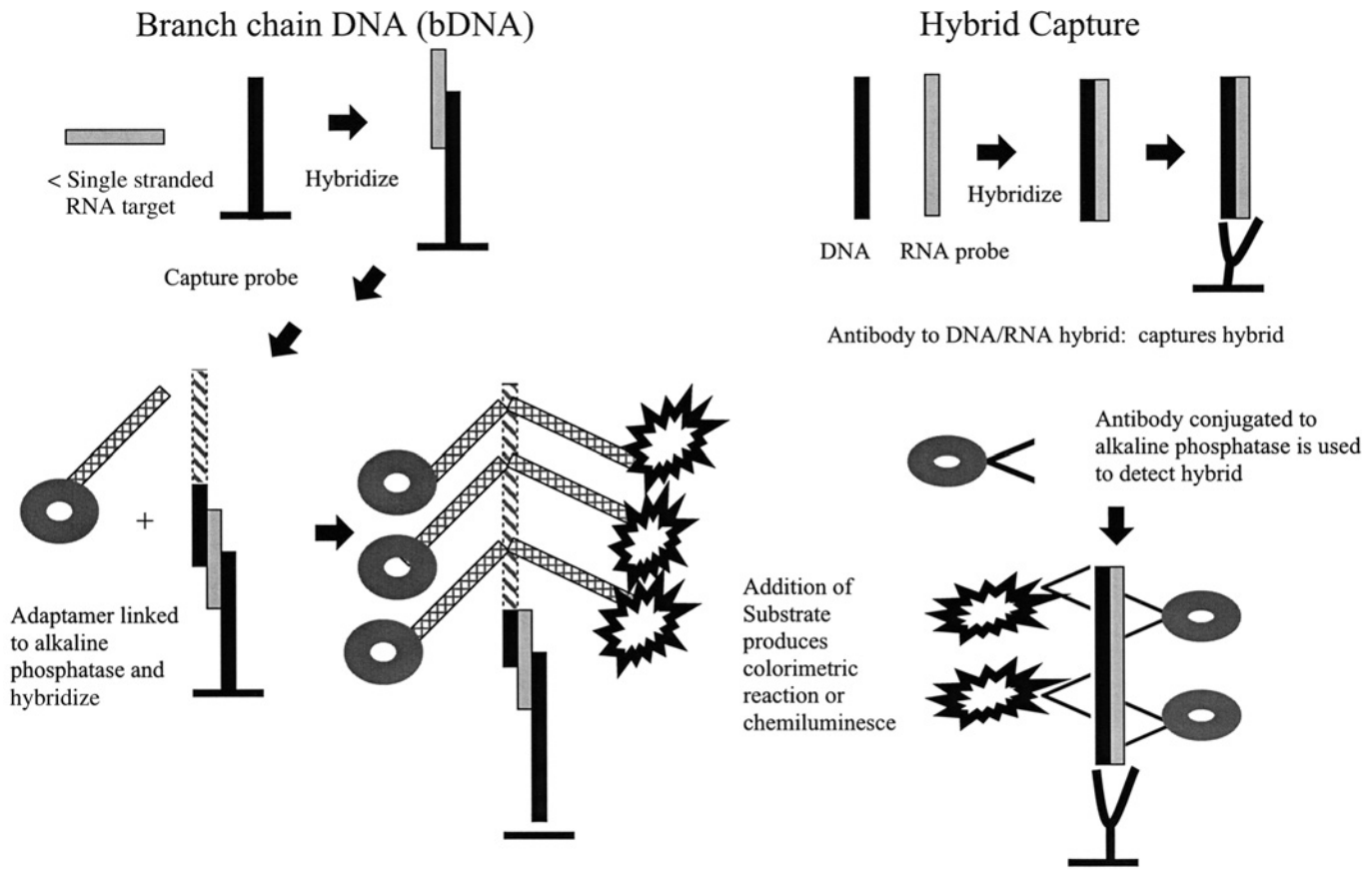


Fig. 6. Nonenzymatic signal amplification.

high or low risk for oncogenic potential. Full-length probes negate the potential for false negatives because SNPs do not inhibit hybridization. The hybrid consists of the viral DNA with the RNA probe. An antibody specific to the DNA/RNA hybrid that is attached to the microtiter plate well wall captures this hybrid. Following washing steps, only the hybrid is retained. Another DNA/RNA-specific antibody that is conjugated to alkaline phosphatase detects the captured hybrid following addition of a chemiluminescence substrate. A luminometer measures the light produced from the reaction. An advantage to this particular system is that thin prep and autocyte specimens obtained during routine pelvic examination can be used later to identify infectious agents in specimens determined by cytopathologists to be suspicious or indeterminate *(29–31)*. For example, many Pap smears are not clearly normal or dysplastic/neoplastic, but are designated ASCUS (atypical squamous cells of unknown significance). Without viral typing, the clinical significance of ASCUS is not clear. Using the hybrid capture assay, however, a reading of ASCUS in the absence of HPV DNA or with HPV DNA from a nononcogenic type would be viewed as benign *(32)*. If high-risk HPV DNA is found, however, the patient would be more likely treated/followed as if her cytopathology results were dysplastic. Unfortunately, there are no universal kits available for "home-brew" assays using the Digene system. However, using currently available reagents and user-supplied probes, a "home-brew" assay can be developed as is illustrated by the development of an in-house Parvovirus B19 test *(33)*.

3.1.2. Branched Chain DNA Clinical assays from Bayer Diagnostics utilize branch chain DNA (bDNA) technologies to detect HIV and HCV using numerous subtype-specific probes to capture the target molecule (Fig. 6B) *(34,35)*. The use of numerous subtype-specific probes negates genetic variations that could result in false-negative reactions or inaccurate quantification, as has been observed when using various target amplification methods *(36)*. Synthetic oligomers that are complementary to the target molecule and to an adaptamer are then hybridized to the captured target. Adaptamers are then hybridized to this bridging molecule and these are extended in a branch formation using additional alkaline-conjugated oligomers. All of these reactions are carried out using different hybridization conditions to ensure appropriate branching reactions from the initial target. Ultimately, a substrate is added to generate a detectable reaction product. The QuantiGene kit can be adapted by the user to develop "home-brew" assays. The limiting step for setting up "home-brew" tests using this system is that the equipment for the various incubations needs to be purchased and is not readily available in most clinical laboratories.

3.1.3. *In Situ* Hybridization *In situ* hybridization is the only method that allows the identification of infectious agents in the context of the tissue's morphology. Although both target and signal amplification methods can be used for *in situ* hybridization, signal amplification using the tyramide

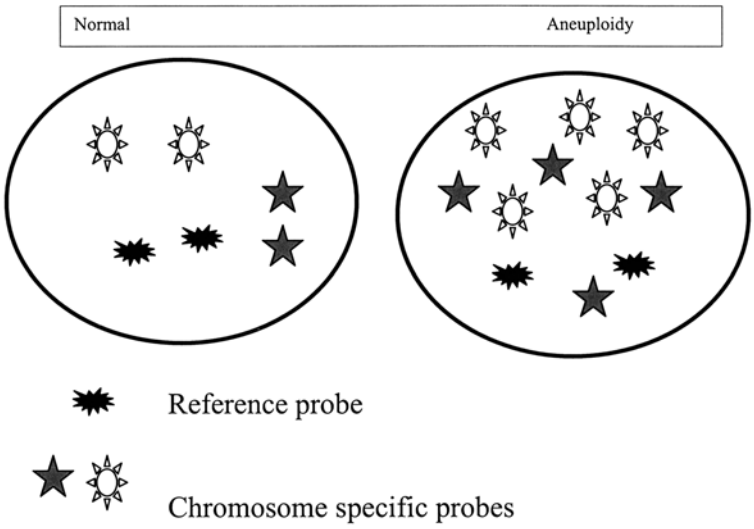


Fig. 7. *In situ* hybridization for aneuploidy.

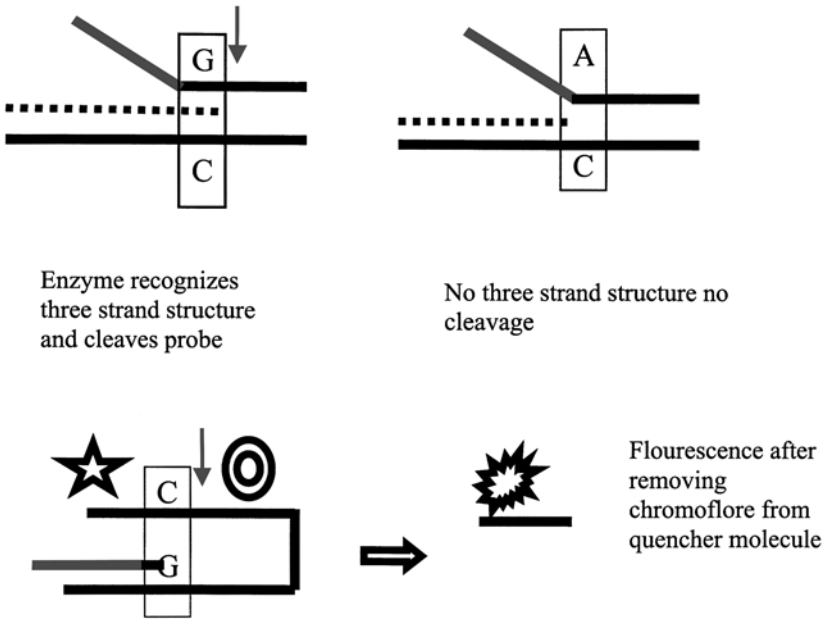


Fig. 8. Enzymatic signal amplification–digestion–cleavase.

amplification method and even bDNA applications are common *(37–39)*. Epstein–Barr virus has been detected in subcutaneous panniculitis and in hypersensitivity reactions to mosquito bites and has been specifically localized to the infiltrating lymphocytes *(40,41)*. The use of *in situ* hybridization was also useful in definitively identifying HPV-containing cells adjacent to molluscum contagiosum bodies *(42)*. Localizing infectious agents to the site of inflammation provides pathologists with the context of the infection that no other method allows. Often infectious agents can contaminant the lesion without actually causing the infection. This caveat is addressed by *in situ* hybridization.

Additionally, different fluorescent probes specific to various chromosomes can also be used to detect aneuploidy (Fig. 7) in cases where cancer is suspected. In this case, different types of "stars" represent the number of different probes. In the normal example, the number of each star is two. Because the normalized probe or star is two, then the ratio is 1 (2 : 2) for each chromosomal probe. However in the test case, this ratio is greater than 1, which would indicate aneuploidy.

3.2. ENZYMATIC-BASED SIGNAL AMPLIFICATION

Another category of signal amplification assays utilizes enzymatic reactions that can detect conformation changes in a target. The most frequently used method in this category utilizes the Cleavase enzyme, which can detect single-basepair changes in the Invader assay (Third Wave Technologies, Madison, WI) *(43,44)* using linear probes, Whereas another method (RAM) combines RCA and strand displacement amplification to amplify a circular DNA probe that serves as the reporter molecule *(45)*.

3.2.1. Cleavage Reactions Flap enzymatic reactions utilize enzymes that can detect changes in the secondary structure of hybridized molecules (Fig. 8). In Fig. 8, the invader oligomer overlaps a single base of interest in the target.

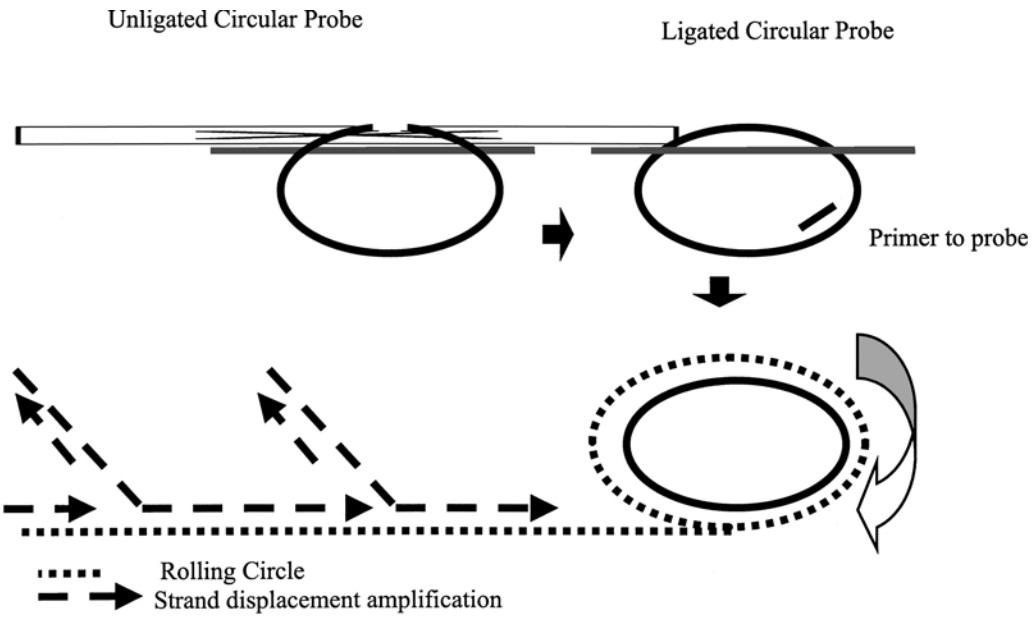


Fig. 9. Enzymatic signal amplification ramification amplification.

A specific probe that is complementary base-pairs with the target's base of interest. The combination of these three molecules forms a structure that is recognized by the cleavase enzyme with then cleaves the primary probe. These cleaved probes then base-pair to the detection molecule, which has a flourochrome and flourochrome-quencher bordering the single base of interest. Again, the structure of the three strands of molecules forms a substrate for the cleavase enzyme. The enzyme cleaves the secondary molecule and releases the flourochrome, which then is detected. None of these reactions takes place if the primary molecule does not bind to its complementary base. Different flourochromes can be used to detect different bases.

3.2.2. Ramification Amplification Ramification amplification is also known as hyperbranched rolling circle or cascade rolling circle amplification *(20,24)*. As its synonyms suggest, this method is similar to the rolling circle amplification with the exception that a ligase closes a circular probe, which then serves as the template for subsequent amplification using reverse and forward primers. As a result, a linear target can be detected using this methodology (Fig. 9).

4. CONCLUSIONS

4.1. SPECIFICITY VS SENSITIVITY Developing testing methodologies that are sensitive and specific can be difficult. Linear primers or probes complementary toward specific areas of the nucleic acid target can result in false-negative results if the target region is polymorphic. These polymorphisms obstruct base-pairing with the primer or probe. Inaccurate quantification resulting from these variations has been demonstrated from studies with human immunodeficiency virus (HIV) *(46)*. These variants can occur naturally or be introduced by enzymes with poor fidelity during the amplification process. Incorporation of erroneous bases during the in vitro replication process can hinder end-point probe-based detection if the misincorporated base is introduced early (within the first three cycles) of the reaction. In this case, the resulting amplicon will not reflect the actual target sequence. If the error is present in the sequence complementary to the probe, the probe might not hybridize because of this mismatch. Methodologies that use low annealing temperatures (i.e., 42°C) might nonspecifically bind to related sequences. To address these issues, methods often incorporate a secondary probe detection step to increase specificity.

Targeting conserved regions of viral, fungal, or bacterial genomes allows the identification of novel unculturable infectious entities and results in higher sensitivity. In bacteria, the DNA sequence that encodes the 16s rRNA has regions that are highly conserved and is referred to as ribosomal DNA (rDNA). Using this approach, different bacterial strains or biotypes have been identified for *Streptococcus* sp., *Mycobacterium* sp., coryneform bacterial isolates *(47,48)* as well as other bacterial species *(49)*. The rDNA sequence most amenable to identification by amplification and subsequent sequencing for fungi is the large subunit. Strain identification of the fungi-nail-tissue-associated *Trichophyton rubrum* and the sinus-cavity-associated *Schizophyllum commune* have been assisted with the aid of conserved primers, followed by sequencing *(50,51)*. Using this approach, *Candida dubliniensis*, which is often incorrectly identified as *Candidia albicans*, has been classified as an emerging yeast pathogen *(52)*. Sequencing need not always be required for the identification of fungal isolates. The information gained through sequencing can be adapted to a more simplistic approach. A line probe assay (LiPA), in which unique probes are affixed to a solid support and hybridized with a labeled product, has been used to identify *C. albicans, C. parapsilosis, C. glabrata, C. tropicalis, C. krusei, C. dubliniensis, Cryptococccus neoformans, Aspergillus fumigatus, A. versicolor, A. nidulans*, and *A. flavus (53)*.

Similarly, primers designed in this manner can detect over 80 different types of HPV *(54,55)*. Although most interest in HPV-associated cancers have classically been for the types associated with cervical dysplasia, identifying cutaneous types

associated with epidermodyplasia verruciformis is useful for confirming the clinical diagnosis. In addition, this approach can be used to detect various herpes viruses associated with lymphomatoid papulosis *(36)*. Although these approaches cited here use PCR, they underscore the importance of primer design and are amenable to other amplification methods utilizing the alternative methods described previously.

4.2. PROBLEMS WITH ENZYMATIC AMPLIFICATION ASSAYS A caveat with all amplification assays is the use of enzymes to facilitate the increase in the target. Inhibitors are often identified in clinical specimens. Internal standards are necessary to rule out the presence of inhibitors that could lead to false-negative results *(56)*. Even in cases of nonprotein-based enzymatic activity (such as in hammerhead ribozymes), various components can result in enzymatic inhibition *(57,58)*.

4.3. CONTAMINATION The fear of amplicon contamination figures prominently in molecular laboratories utilizing amplification methods. Because small amounts of amplicons can result in widespread contamination, clinical laboratories must implement numerous contamination control protocols such as dedicated pipettors, aerosol-resistant tips, different work areas for master mix preparation, specimen extraction and amplicon detection, and unidirectional workflows. Some commercial kits contain reagents that control contamination by making the amplicon sensitive to enzymatic degradation (amperase; Roche Molecular Systems, Indiannapolis, IN). Closed-tube systems decrease the risk for contamination by physically separating the amplified product from the environment; however, precautions must still be made to avoid inadvertently contaminating the reagent preparation area and specimen extraction area with positive specimens or control material.

4.4. FLEXIBILITY Commercially available kits are not always available for esoteric targets. For these less common agents, the clinical laboratory must develop and validate in-house molecular assays for these tests. User-developed assays for esoteric testing most often use PCR technology. One possible explanation is the higher prevalence of PCR equipment and experience in clinical laboratories. Yet another explanation could be relative ease of optimizing this assay as the higher specificity associated with high-temperature annealing *(59,60,61)*. As other approaches become more frequently used, it is anticipated that users will learn how to modify these platforms to address esoteric testing.

Finally, molecular methods are increasingly being used to assist in the diagnosis of various genetic and infectious diseases. The use of these methods improves patient care by decreasing the turnaround times (relative to culture or serology) and for confirming diagnoses based on clinical observations. Technology is rapidly advancing to decrease the time required for these assays as well as their costs. Many of the methods discussed in this chapter could facilitate the implementation of automated testing in molecular diagnostic laboratories. In the future, the range of molecular pathology will extend beyond nucleic-acid-based detection systems and will evolve to utilize information derived from the proteome as well to provide information on optimizing treatment and predicting outcomes for patients with both genetic and infectious diseases.

REFERENCES

1. Oksi, J., et al. Erythema migrans—influence of posture. Case report. *APMIS* **108:**649–650, 2000.
2. Coleme-Grimmer, M. I., et al. *Borrelia burgdorferi* and *Borrelia hermsii* DNA are not associated with morphea or lichen sclerosus and atrophicus in the southwestern. *Arch. Dermatol.* **133:**1174, 1997.
3. Burdick, A. E., et al. Hansen's disease in a patient with a history of sarcoidosis. *Int. J. Lepr. Other Mycobact. Dis.* **68:**307–311, 2000.
4. Donoghue, H. D., et al. PCR primers that can detect low levels of Mycobacterium leprae DNA. *J. Med. Microbiol.* **50:**177–182, 2001.
5. Soini, H. M. J. Molecular diagnosis of mycobacteria. *Clin. Chem.* **47:**809–814, 2001.
6. Yen, A. R. P., Cortes-Franco, R., and Tyring, S. K. Detection of *Mycobacterium tuberculosis* in erythema induratum of Bazin using polymerase chain reaction. *Arch. Dermatol*. **133:**532–533, 1997.
7. Abalos, F. M., et al. *Mycobacterium ulcerans* infection (Buruli ulcer): a case report of the disseminated nonulcerative form. *Ann. Diagn. Pathol*. **4:**386–390, 2000.
8. Yamanaka, K., et al. Severe dissseminated BCG infection in an 8 year old girl. *Nagoya J. Med. Sci.* **63:**123–128, 2000.
9. Dean, F. B., et al. Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. USA* **99(8):**5261–5266, 2002.
10. de Barbeyrac, B., et al. Detection of *Chlamydia trachomatis* in symptomatic and asymptomatic populations with urogenital specimens by AMP CT (Gen-probe incorporated) compared to others commercially available amplification assays. *Diagn. Microbiol. Infect. Dis.* **37(3):**181–185, 2000.
11. Linnen, J. M., et al. Sensitive detection of genetic variants of HIV-1 and HCV with an HIV-1/HCV assay based on transcription-mediated amplification. *J. Virol. Methods* **102(1–2):**139–155, 2002.
12. Akduman, D., et al. Evaluation of a strand displacement amplification assay (BD ProbeTec-SDA) for detection of *Neisseria gonorrhoeae* in urine specimens. *J. Clin. Microbiol.* **40(1):**281–283, 2002.
13. Barrett, A., Magee, J. G., and Freeman, R. An evaluation of the BD ProbeTec ET system for the direct detection of *Mycobacterium tuberculosis* in respiratory samples. *J. Med. Microbiol.* **51(10):**895–898, 2002.
14. Walter, N. G. and Strunk, G. Strand displacement amplification as an in vitro model for rolling-circle replication: deletion formation and evolution during serial transfer. *Proc. Natl. Acad. Sci. USA* **91(17):**7937–7941, 1994.
15. Baner, J., et al. Signal amplification of padlock probes by rolling circle replication. *Nucleic Acids Res.* **26(22):**5073–5078, 1998.
16. Christian, A. T., et al. Detection of DNA point mutations and mRNA expression levels by rolling circle amplification in individual cells. *Proc. Natl. Acad. Sci. USA* **98(25):**14238–14243, 2001.
17. Demidov, V. V. Rolling-circle amplification in DNA diagnostics: the power of simplicity. *Expert Rev. Mol. Diagn.* **2(6):**542–548, 2002.
18. Faruqi, A. F., et al. High-throughput genotyping of single nucleotide polymorphisms with rolling circle amplification. *BMC Genomics* **2(1):**4, 2001.
19. Hatch, A., et al. Rolling circle amplification of DNA immobilized on solid surfaces and its application to multiplex mutation detection. *Genet. Anal.* **15(2):**35–40, 1999.
20. Lizardi, P. M., et al. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat. Genet.* **19(3):**225–232, 1998.
21. Nallur, G., et al. Signal amplification by rolling circle amplification on DNA microarrays. *Nucleic Acids Res*. **29(23):**E118, 2001.
22. Qi, X., et al. L–RCA (ligation–rolling circle amplification): a general method for genotyping of single nucleotide polymorphisms (SNPs). *Nucleic Acids Res.* **29(22):**E116, 2001.
23. Rosler, A., et al. Rolling circle amplification for scoring single nucleotide polymorphisms. *Nucleosides Nucleotides Nucleic Acids* **20(4–7):**893–894, 2001.
24. Thomas, D. C., Nardone, G. A., and Randall S. K. Amplification of padlock probes for DNA diagnostics by cascade rolling circle amplification or the polymerase chain reaction. *Arch. Pathol. Lab. Med.* **123(12):**1170–1176, 1999.

25. Wu, H. C., et al. DNA sequencing using rolling circle amplification and precision glass syringes in a high-throughput liquid handling system. *Biotechniques* **34(1):**204–207, 2003.
26. van Doornum, G. J., et al. Comparison between the LCx Probe system and the COBAS AMPLICOR system for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* infections in patients attending a clinic for treatment of sexually transmitted diseases in Amsterdam, The Netherlands. *J. Clin. Microbiol*. **39(3):**829– 835, 2001.
27. Winter, A. J., et al. Comparison of a ligase chain reaction-based assay and cell culture for detection of pharyngeal carriage of *Chlamydia trachomatis*. *J. Clin. Microbiol.* **38(9):**3502–3504, 2000.
28. Reyes, A. A., et al. Linked Linear amplification (LLA): a new method for the amplification of DNA. *Clin. Chem.* **47:**31–40, 2001.
29. Manos, M. M., et al. Identifying women with cervical neoplasia. *JAMA* **281:**1605–1647, 1999.
30. Cullen, A. P., et al. Rapid detection and typing of herpes simplex DNA in clinical specimens by the hybrid capture II signal amplification probe test. *J. Clin. Micro. biol.* **35:**2275–2278, 1997.
31. Hong, I. S., et al. Comparative analysis of a liquid-based Pap test and concurrent HPV DNA assay of residual samples. A study of 608 cases. *Acta. Cytol.* **46(5):**828–834, 2002.
32. Khanna, N., et al. Human papillomavirus absence predicts normal cervical histopathologic findings with abnormal papanicolaou smears: a study of a university-based inner city population. *J. Hum. Virol*. **4(5):**283–287, 2001.
33. Boggino, H., et al. A quantitative probe method for the detection of Parvovirus DNA. *J. Clin. Lab. Analy.* **14(1):**38–41, 2000.
34. Gleaves, C. A., et al. Multicenter evaluation of the Bayer VERSANT HIV-1 RNA 3.0 assay: analytical and clinical performance. *J. Clin. Virol.* **25(2):**205–216, 2002.
35. Trimoulet, P., et al. Evaluation of the VERSANT HCV RNA 3.0 assay for quantification of hepatitis C virus RNA in serum. *J. Clin. Microbiol.* **40(6):**2031–2036, 2002.
36. Parekh, B., et al. Impact of HIV type 1 subtype variation on Viral RNA quantitation. Aids research and human retroviruses. **15(2):**33–142, 1999.
37. Evans, M. F., et al. Biotinyl-tyramide-based in situ hybridization signal patterns distinguish human papillomavirus type and grade of cervical intraepithelial neoplasia. *Mod. Pathol.* **15(12):**339–347, 2002.
38. Kenny, D., Shen, L. P., and Kolberg, J.A. Detection of viral infection and gene expression in clinical tissue specimens using branched DNA (bDNA) in situ hybridization. *J. Histochem. Cytochem*. **50(9):**1219–1227, 2002.
39. Van Tine, B. A., et al. In situ analysis of the transcriptional activity of integrated viral DNA using tyramide–FISH. *Dev. Biol.* (*Basel*) **106:**381–385, 2001.
40. Ishihara, S., et al. Hypersensitivity in mosquito bites is not an allergic disease but an Epstein Barr virus associated lymphoproliferative disease. *Int. J. Hematol.* **72:**223–228, 2000.
41. Abe, Y., et al. Subcutaneous panniculitis by Epstein Barr virus infected natural killer (NK) cell proliferation terminating in aggressive subcutaneous NK cell lymphoma. *Am. J. Hematol.* **64(3):** 221–225, 2000.
42. Payne, D. A., et al. Coexistance of Molluscum contagiosum with HPV in the same lesion. *J. Am. Acad. Dermatol.* **36(4):**641–644, 1997.
43. Nevilie, M., et al. Characterization of cytochrome P450 2D6 alleles using the Invader system. *Biotechniques* **Suppl:**34–43, 2002.
44. Evans, J. G., and Lee-Tataseo, C. Determination of the factor V Leiden single-nucleotide polymorphism in a commercial clinical laboratory by use of NanoChip microelectronic array technology. *Clin. Chem*. **48(9):**1406–1411, 2002.
45. Zhang, W., et al. Detection of *Chlamydia trachomatis by isothermal ramification amplification method: a feasibility study. J. Clin. Microbiol.* **40(1):**128–132, 2002.
46. Lazar, J. G., et al. Hybrid capture: a sensitive signal amplification based chemiluminescence test for the detection and quantification of human viral and bacterial pathogens. *J. Clin. Ligand Assay* **22:**139–151, 1999.
47. Tang, Y. W., et al. Identification of coryneform bacterial isolates by ribosomal DNA sequences. *J. Clin. Microbiol.* **38(4):**1676–1678, 2000.
48. Patel, J. B., et al.Sequence based identification of Mycobacterium species using the Microseq rDNA bacterial identification system. *J. Clin. Microbiol.* **38(1):**246–251, 2000.
49. Lundberg, F., et al. External ventricular drainage catheters: effect of surface heparinization on bacterial colonization and infection. *Acta Neurochir*. **142:**1377–1383, 2000.
50. Jackson, C. J., et al. Strain identification of *Trichophyton rubrum* by specific amplification of subrepeat elements in ther ribosomal DNA nontranscribed spacer. *J. Clin. Microbiol*. **38(12):** 4527–4534, 2000.
51. Buzina, W., et al. Development of molecular methods for identification of Schizophyllum commune form clinical samples. *J. Clin. Microbiol.* **39(7):**2391–2396, 2001.
52. Kurzai, O., et al. Molecular and phenotypic identification of the yeast pathogen *Candida dubliniensis. J. Mol. Med.* **78(9):**521–529, 2000.
53. Martin, C., et al. Development of a PCR based line probe assay for identification of fungal pathogens. *J. Clin. Microbiol.* **38(10):**3735–3742, 2000.
54. Manos, M. M., et al.The use of polymerase chain reaction amplification for the detection of genital human papillomaviruses. *Cancer Cells* **7:**209–214, 1989.
55. Kempf, W., et al. Lymphomatoid papulosis and human herpesviruses—a PCR based evaluation for the presence of human herpesvirus 6, 7, 8 and related herpesviruses. *J Cutan. Pathol*. **28:** 29–33, 2001.
56. Bezold, G., et al. Detection of herpes simplex virus and varicella-zoster virus in clinical swabs: frequent inhibition of PCR as determined by internal controls. *Mol. Diagn.* **5:**279–284, 2000.
57. Stage, T. K., Hertel, K. J., and Uhlenbeck, O. C. Inhibition of the hammerhead ribozyme by neomycin. *RNA* **1(1):**95–101, 1995.
58. Jenne, A., et al. Rapid identification and characterization of hammerhead-ribozyme inhibitors using fluorescence-based technology. *Nat. Biotechnol.* **19(1):**56–61, 2001.
59. Payne D, et al. Increased buffer pH enhances sensitivity and specificity of human papillomavirus detection using consensus primer based PCR. *J. Virol. Methods* **52:**105–110, 1995.
60. Yang, S., et al. Development of a rapid detection assay for bacterial identification and speciation. *Acad. Emerg. Med.* **8:**494, 2001.
61. Schweitzer, B., et al. Combining nucleic acid amplification and detection. *Curr. Opin. Biotechno.* **12:**21–27, 2001.

9 Electrophoretic Methods for Mutation Detection and DNA Sequencing

W. EDWARD HIGHSMITH, JR.

1. INTRODUCTION

Since its development in 1985, polymerase chain reaction (PCR) has revolutionized basic and applied research *(1,2)*. In 1993, Mullis was awarded the Nobel Prize in Chemistry for the development of PCR. With DNA or cDNA as a template, millions of copies of a target sequence are generated during the reaction. Introduction of the thermophilic *Thermus aquaticus* polymerase increased the specificity of the reaction and made automation and routine use possible *(3–5)*. The ability of PCR to produce multiple copies of a discrete portion of the genome has resulted in its incorporation into techniques used in a wide variety of research and clinical applications. An extraordinary range of clinical applications of PCR have emerged, including diagnosis of inherited disease, human leukocyte antigen (HLA) typing, identity testing, infectious disease diagnosis and management, hematologic disease diagnosis and staging, and susceptibility testing for cancer.

The development of technically simple and reliable methods to detect sequence variations in specific genes is becoming more important as the number of genes associated with specific diseases grows. DNA sequencing is considered the "gold standard" for characterization of specific nucleotide alteration(s) that result in genetic disease. Although sequencing was long considered too cumbersome, expensive, and operator dependent for use in the clinical laboratory, a combination of clinical need and improved technology has brought automated DNA sequencing into routine clinical use. However, even though sequencing technology is now firmly entrenched in the clinical molecular diagnostics laboratory, it is still too expensive and time-consuming for all the laboratory's mutation-detection needs. There are a number of PCR-based mutation-detection strategies that can be used to identify both characterized and uncharacterized mutations and sequence variations.

The degree of allelic heterogeneity, or the number of different mutations in a single gene (each of which cause a specific disorder), influences the method used for mutation detection. For diseases that exhibit no or limited heterogeneity (e.g., sickle cell anemia), assay systems designed to detect specific mutations are appropriate. These types of strategy are also appropriate for disorders in which allelic heterogeneity is high, but only a limited set of mutations are typically analyzed, such as cystic fibrosis. For disorders in which the mutational spectrum is wide (e.g., Duchenne/Becker muscular dystrophy or multiple endocrine neoplasia), a scanning method is needed. A scanning method is also appropriate for analysis of newly identified disease genes, for which there is little or no information regarding the number of disease-causing mutations.

In most applications, PCR is used to amplify specific regions of DNA known to carry or suspected of carrying a mutation. The specific DNA sequence, whether normal or mutated, is then identified by hybridization or electrophoretic separation of the PCR products. In a few techniques, PCR itself is designed to specifically identify the normal and mutant DNA sequence. In this chapter, we will discuss electrophoretic PCR-based techniques for the analysis of DNA and RNA in the clinical laboratory.

2. POLYMERASE CHAIN REACTION

The standard PCR reaction takes place in a closed tube that is subject to a series of temperature changes or cycles. The initial phase of a cycle involves heating the reaction mixture, including template, primers, polymerase, and free deoxynucleotides, to 92–96°C to denature the double-stranded template and, in subsequent cycles, the double-stranded reaction products. When the reaction mixture is cooled during the second phase of the cycle, oligonucleotide primers anneal to opposite strands of the template flanking the target sequence. The optimal annealing temperature ranges from 45°C to 72°C and is related to the melting temperature of the primers, which is determined by their length and G + C content, and the concentration of monovalent cations. In the final stage, at 72°C, the polymerase extends the primers from the 3′ ends toward each other along the template strands. The entire cycle is ordinarily repeated 25–40 times, and because the products are also used as template in subsequent reactions, there is an exponential increase in the number of copies, resulting in the generation of a discrete portion of DNA, which is defined by the 5′ ends of the primers.

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

Table 1
Equations Describing Electrophoretic Driving Force

Eq. (1)—Electrophoretic driving force

$$F = XQ = \frac{VQ}{l} \quad (1)$$

where F is the force driving the ion forward, X is the electric field strength or voltage drop, and V is the applied voltage, and l is the length of the gel.

Eq. (2)—Stokes Law

$$F' = 6\pi r\eta v \quad (2)$$

where F' is the counterforce, π is a constant (3.14159), r is the radius of the analyte, η is the viscosity of the matrix, η and, v is the velocity of the analyte movement (l/t)

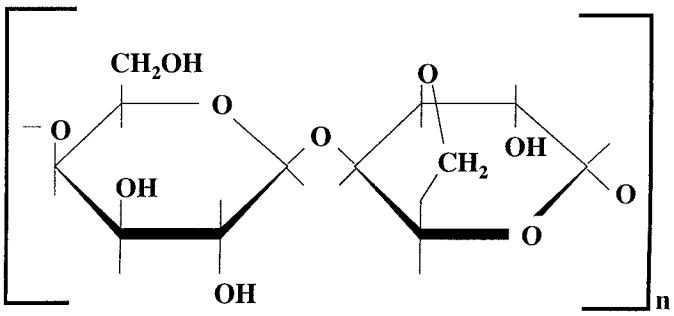


Fig. 1. Agarbiose molecule. Structure of agarbiose, the repeating subunit of agarose.

3. FUNDAMENTALS OF ELECTROPHORESIS

Electrophoresis refers to the migration of charged molecules through a liquid or gel medium under the influence of an electric field. Zone electrophoresis, or the migration of macromolecules through a porous support medium, or gel, is almost universally used in molecular biology laboratories today. Electrophoresis (all discussion in this chapter will involve zone electrophoresis but will be referred to as electrophoresis for brevity) is a powerful separation tool, being able to detect the differential migration of macromolecules with only subtly different structures. The two formats that are used are slab gel electrophoresis, in which the electrophoresis takes place in support medium, agarose, or polyacrylamide, and capillary electrophoresis (CE), an instrumental technique in which the electrophoresis takes place in a capillary tube.

The rate of migration of a macromolecule through a gel matrix is dependent on several factors, including (1) the net charge on the molecule at the pH at which the assay is conducted, (2) the size and shape of the molecule, (3) the electric field strength or voltage drop, (4) the pore size of the gel, and (5) temperature. The forces acting on the analyte to drive it through the gel are the charge on the molecule and the electric field strength. Equation (1) describes the electrophoretic driving force (Table 1).

The forces acting to retard the movement of the molecule are the frictional forces, determined by the velocity of the analyte, the pore size of the gel, and the size and shape of the molecule. The opposition of the acceleration of the analyte by the frictional forces is described by Stokes' Law as shown in Eq. (2) (Table 1). When the electrophoretic driving force equals the frictional force ($F = F'$), the result is a constant velocity of the analyte molecule through the gel matrix [Eq. (2), Table 1]. The term v/X describes the velocity of the analyte through the gel matrix (cm/s) per unit field strength (V/cm) under constant conditions (i.e., the same buffer conditions and the same gel viscosity). This term (given by the symbol μ) is defined as the electrophoretic mobility of the analyte (expressed as cm^2/V-s). From the units of the electrophoretic mobility, it can be seen that if the gel is run under constant-voltage conditions (V/cm constant) for a given period of time (s), then (cm^2/V-s)(V-s/cm) is distance of migration (in cm). Thus, if gels are run such that the product of the voltage and the time are constant, the same analyte will migrate the same distance into the gel. For convenience, this is typically expressed as volt-hours. For example, if one gel is run at 50 V for 10 h (500 V-h) and an identical gel (constant length, viscosity, and buffer concentration) is run at 100 V for 5 h (500 V-h), the same analyte will appear at the same position on both gels. This ability to reproduce gel profiles is one of the principal reasons that constant-voltage run conditions are preferred for DNA analysis. An exception to this is the preferred conditions for DNA sequencing. As will be discussed later, sequencing gels are run at elevated temperature to ensure the adequate denaturation of the single-stranded DNA molecules; in this case, running the gel at constant watts is helpful in maintaining an even heating of the gel.

In protein analysis, the pH of the electrophoresis buffer can be a powerful tool in optimizing specific separations. This is because the type (acidic or basic) and number of ionizable groups found on proteins are variable. However, for DNA analysis, it is the charge of the phosphate backbone that is dominant. Therefore, DNA electrophoresis is typically performed at a slightly alkaline pH to ensure full ionization of the phosphate residues.

3.1. THE GEL MATRIX: AGAROSE There are two types of gel matrix in common use in DNA laboratories: agarose and polyacrylamide. Agarose is a polysaccharide commercially derived from seaweed. The agarose polymer consists of multiple agarbiose molecules linked together into linear chains, with an average molecular weight of 120,000 daltons. The agarbiose subunit is a disaccharide consisting of β-d-galactose and 3,6-anhydro-α-l-galactose (Fig. 1) *(6)*. The partially purified material, agar, consists of noncharged polymer chains, agarose, and negatively charged chains. The negative charges are typically the result of sulfate ($-SO_4$) residues. In general, the more highly purified the agarose, the lower the sulfate concentration, the higher the quality of the separation, and the higher the price. Agarose is supplied as a white, nonhydroscopic powder. A gel is prepared by mixing agarose powder with buffer, boiling the mixture, pouring the molten gel into a casting tray, and cooling. During this process, the agarose chains shift from existing in solution as random coils to a structure in which the chains are bundled into double helices. The average pore size for agarose gels are typically in the range 100–300 nm^3. Agarose gels are used at concentrations near 1% (w/v) for separating DNA fragments in the 1- to 20-kb size range. Examples of applications common in the DNA diagnostic laboratory include restriction digestion analysis of large plasmids and Southern transfer analysis of genomic DNA.

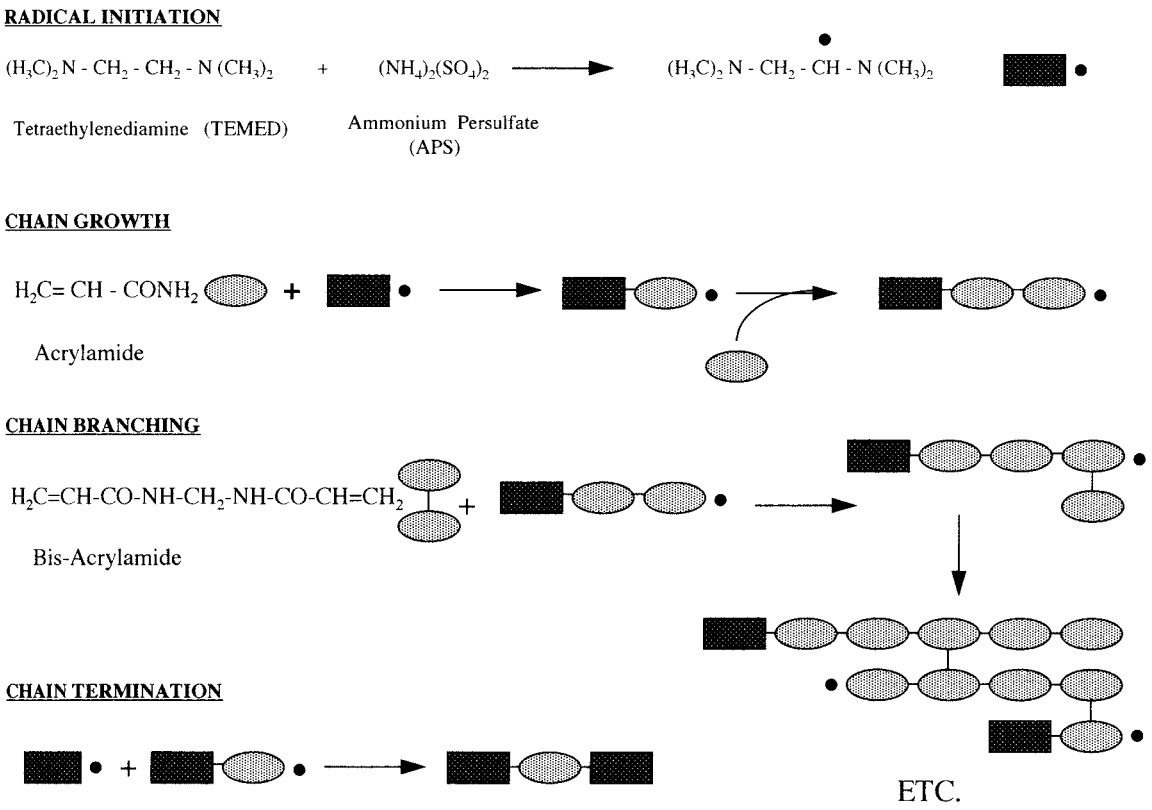


Fig. 2. Schematic demonstrating the polymerization of acrylamide and bis-acrylamide monomers.

A series of chemically modified agaroses with smaller pore sizes were developed by scientists at FMC BioProducts. These modified gel matrices are prepared by melting and casting similarly to unmodified agarose, but because of the smaller pore sizes, they are very useful for DNA separations in the 100- to 1000-bp size range and are widely used in molecular diagnostics. These products are marketed under the trade names NuSieve® and MetaPhor® agaroses (Cambrex Bio Science Walkersville, Inc., Walkersville, MD).

3.2. THE GEL MATRIX: POLYACRYLAMIDE The advent of PCR has had profound effects on the clinical laboratory's ability to use the tools of molecular biology for clinical diagnostic testing. However, it is important to understand the effect that PCR technology has had on the availability and choice of electrophoretic techniques in the clinical laboratory. The principal effect of PCR on the practice of electrophoresis has been to shift the size of the analytes from large fragments of DNA, as used in the Southern transfer, to small fragments of DNA, typically from 100 to 1000 bp in length. Although the chemically modified agaroses (which have been optimized for the separation of low-molecular-weight DNA fragments) are the most commonly used matrixes for the analysis of PCR products, it is necessary to use polyacrylamide gels for very high resolution.

Polyacrylamide gels are prepared from a monomer, acrylamide, and a crosslinker, typically bis-acrylamide. (*Note*: Acrylamide is a potent neurotoxin and is readily absorbed through the skin. When using acrylamide in aqueous solution, wear gloves and a lab coat. When weighing powdered acrylamide, wear goggles and perform the weighing operation in a chemical hood.) The first step in preparing a polyacrylamide gel is to add a free-radical initiator to a solution of monomer and crosslinker. The most commonly used initiator system is tetra ethylene diamine (TEMED) and ammonium persulfate (APS) (Fig. 2). APS reacts with the TEMED to form a TEMED derivative with a free, or unpaired, electron. This type of molecule is termed a free radical and is highly reactive. The TEMED radical reacts with an acrylamide molecule, forming a TEMED–acrylamide radical. This first step of the polymerization process is termed "chain initiation." The next steps are chain elongation, in which the polymer chain grows by repetitive addition of acrylamide monomers to the growing chain with the free-radical terminus. Chain branching occurs when a bis-acrylamide molecule is added to the end of the chain. Chain termination occurs when two free radicals react, giving a stable compound with paired electrons. In order to achieve complete polymerization, it is important that compounds that quench free-radical reactions, such as alcohols or oxygen, be excluded from the reaction mixture. Oxygen is typically removed from the acrylamide/bis-acrylamide solution by degassing under vacuum for 15–30 min prior to the addition of the initiators. To avoid oxygen contact during the polymerization process, the gel is generally cast between two glass plates. After polymerization, the gel is run in a vertical format.

The pore size of the final polyacrylamide gel is determined by the concentration of the acrylamide monomer and the ratio of the crosslinker to monomer. These parameters are referred to as %T and %C, respectively, and are defined as follows: %T = mass of all monomers and crosslinkers per 100 mL volume and

%C = mass of the crosslinker ÷ mass of all monomers and crosslinkers per 100 mL volume. The pore sizes in polyacrylamide gels are typically much smaller than those in agarose gels. Thus, they are used for DNA fragments of smaller size. Gels with %C of approx 3–5% and %T ranging from 5 to 15% are the most often used gels in the DNA laboratory, giving superior separations in the 100- to 1000-bp size range.

A series of gel products with improved performance for specific applications were developed by scientists at AT Biochem (Malvern, PA). These materials use acrylamide as a monomer, but incorporate proprietary comonomers and novel crosslinkers. Recently, this line of products has been acquired Cambrex Bio Science Walkersville, Inc. (Walkersville, MD). A gel matrix offering longer reads for DNA sequencing and a matrix optimized for the detection of conformational changes in single-strand conformation polymorphism or heteroduplex analysis are currently marketed under the trade names Long Ranger® and Mutation Detection Enhancement Gel (or MDE®), respectively.

4. CAPILLARY ELECTROPHORESIS

Capillary electrophoresis (CE) is an extremely powerful instrumental technique that has introduced automation into the clinical molecular laboratory. The method, originally described by Jorgenson and colleagues at the University of North Carolina in 1983 *(7)*, utilizes electrophoresis in the thin, fused silica capillary columns that had been recently developed for use in capillary gas chromatography. Electrophoresis in free solution (without porous matrices) in open tubes was not new, having been described by Tiselius in 1937 *(8)*. In the original format, however, resolution was severely compromised by the heat produced by the passage of the electric current though the solution, as the uneven heating created unwanted mixing within the electrophoresis tube. The introduction of the capillary tube solved the Joule heating problem, as the capillaries used (with inner diameters of 50–100 μm) have a very high surface-to-volume ratio and are extremely efficient at dissipating heat, even at applied voltages of 10,000–20,000 V.

Detection of the separated products is accomplished with detectors similar to those used in high-performance liquid chromotography (HPLC). Because the detector is fixed in position near the end of the capillary column, an analyte must pass through the entire length of the capillary before being detected. This represents a significant advantage in terms of resolution compared to slab gels. In CE, each peak is detected only after having experienced the separating power of the entire column. In contrast, in slab gels, bands obviously must remain on the gel to be detected, with the more difficult to resolve, higher-molecular weight species being the ones that have migrated the least distance into the gel, thereby experiencing the least amount of separation. Because the fused silica capillaries are optically clear (after removal of a small window in the polyimide coating that capillaries are supplied with in order to prevent breakage), detection can take place directly in the column; there is not need for any connective plumbing between the column and the detector. This has significant advantages in terms of both instrumental simplicity and elimination of any band-broadening effects as a result of an imperfect fluid path connections. Although ultraviolet (UV)-visible detectors can be used to monitor DNA separations, the majority of applications utilize laser-induced fluorescence (LIF) detectors. Because DNA has very low intrinsic fluorescence, the DNA must be labeled with a fluorescent tag, typically by amplifying the target with a labeled PCR primer, in order to use an LIF detector. Modern CE-LIF detectors can monitor four of more different emission wavelengths; thus, in addition to being an extremely sensitive detection method, fluorescence has the great advantage of being compatible with multiplexing, or analyzing several different PCR products at once. A generalized schematic of a CE instrument is shown in Fig. 3.

Although CE was rapidly shown to be a powerful separation tools for small molecules, the application to macromolecules, proteins, and nucleic acids required the solution to several fundamental problems. The surface of the silica column is compose of large numbers of silanol groups (–Si–OH). At high pH, these slightly acidic hydroxyls ionize, giving a surface with a high net negative charge. This negative charge attracts cations from the buffer. These cations are bound to the charged silica wall in two identifiable layers, the Inner Helmholtz or Stern Layer, and the Outer Helmholtz Plane (OHP). The inner layer is very tightly bound and is essentially static, but when an electric field is applied, the positive ions of the OHP, along with their water of hydration, migrate to the cathode. As the waters of hydration are hydrogen-bonded to the water molecules of the bulk solution inside the capillary, there is a net flow of liquid, or buffer, from the anode to the cathode. This bulk flow of buffer within the capillary column on application of an electric field is the endo-osmotic flow (EOF) *(9)*. The EOF can be manipulated, typically be alterations in the buffer pH, and is a useful tool in designing separations of mixtures of small molecules that are heterogeneous in structure and charge. For nucleic acids, however, which are uniform in structure and charge, the EOF is highly detrimental to size-based separations. Another problem with separations of macromolecules by CE is the pronounced tendency for these molecules to interact with the column wall, sometimes irreversibly, but always with a drastic and detrimental effect on the peak shapes and separation efficiency. In addition, electrophoresis is typically used to separate compounds that differ from each other by charge or shape. In the case of nucleic acids, two of the terms in Eq. (2), the frictional coefficient and the charge, change together as a function of the size of the nucleic acid fragment (i.e., DNA molecules display a constant frictional coefficient-to-charge ratio). Thus, there is no separation of different sizes of DNA molecules by electrophoresis without a sieving matrix.

Fortunately, these three problems have been solved. The first two problems, elimination of the EOF and adsorption of macromolecules to the column wall, have the same solution—coating the interior wall of the column with a noncharged coating. There have been a variety of column-coating methods utilized, including those borrowed from the capillary gas chromatography field where capillary coatings, with a wide variety of polymers, are used as solid phases for chromatographic separations. Other methods include derivitization of the negatively charged silanol residues on the surface of the silica column with silating reagents.

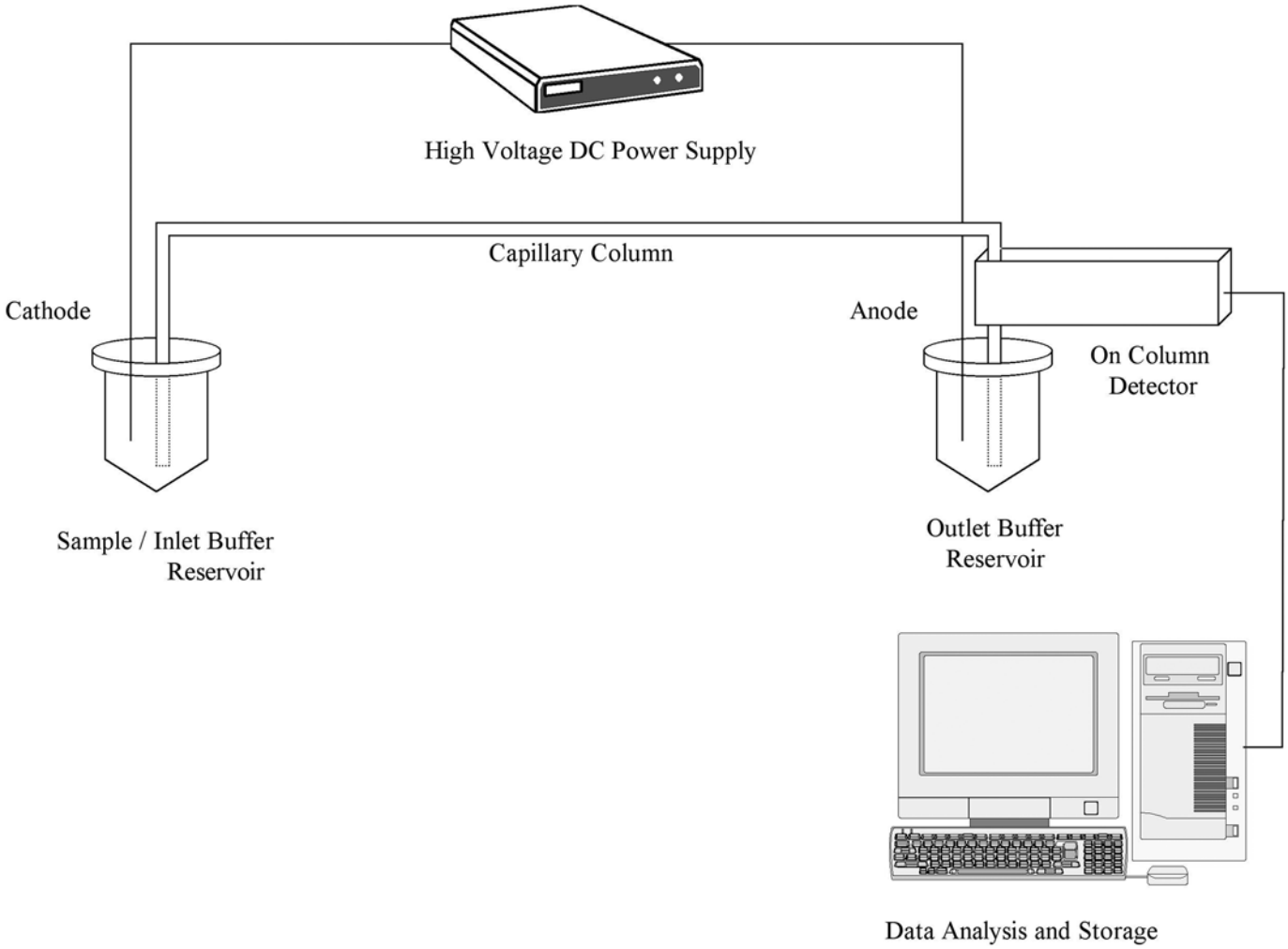


Fig. 3. A schematic of a capillary electrophoresis instrument. The basic components of a CE system are a high-voltage (5–30 kV) power supply, a polyimide-coated, fused-silica capillary column (20–100 μm inner diameter, 20–50 in length) filled with buffer, buffer reservoirs to accommodate both the column and the electrodes, a detector (nucleic acid applications typically utilize a LIF detector), and a desktop computer for instrument programming and data analysis and storage.

The first attempts to introduce a sieving matrix into a capillary column were a variety of methods to polymerize acrylamide inside the column. Although it is possible to prepare a gel-filled column, the technique was plagued with problems, including the formation of bubbles during polymerization, which block the electric current, extrusion of the gel during electrophoresis by the EOF, and limited number of times the column could be used. In 1993, Karger and colleagues at Boston University introduced the notion of replaceable sieving matrices. These matrices consisted of, not a gel of bis-acrylamide crosslinked polyacrylamide, but a viscous solution of non-crosslinked, high-molecular-weight linear polyacrylamide *(10)*. At concentrations greater than the entanglement threshold, these polymer solutions acted in a very similar way to gels in DNA size separations. In gels, the size-based separation is mediated by the pore size. In solutions of water-soluble polymers, the separation is mediated by the pores between entangled polymer chains. This breakthrough enabled the development of commercial, capillary-based sequence analyzers using coated capillaries and linear acrylamide polymer solutions. The systems using coated capillaries and linear polyacrylamide are successful at providing DNA sequencing reads of well over 500 bases and form the basis for several commercial CE-based automated sequence analyzers. However, the coated capillaries are relatively expensive to prepare and have a modest lifetime (approx 200 runs). More recently, a variety of polymers, including polydimethylacrylamide, polyethylene oxide, polyvinylpryolodone, and the recently described polyhydroxyethylacrylamide *(11–13)* have been shown to not only provide sieving-based DNA size separations but also self-coating characteristics that allow them to be used with bare silica capillary columns. The details on the mechanisms of polymer coating of silica surfaces is still somewhat obscure, but apparently they involve both interactions between hydrophobic domains of the polymer and the hydrophobic siloxanes of the silica surface and hydrogen-bonding between the polymer and the surface silanols *(14,15)*. The read length obtainable using a variety of polymers has been investigated and it is clear that linear polyacrylamide is superior in terms of read length. The price, however, is paid in terms of the requirement to use the more

expensive, shorter-lifespan coated columns. Currently, CE systems with DNA analysis capabilities are available from several manufacturers.

5. APPLICATIONS OF ELECTROPHORESIS IN THE DNA DIAGNOSTICS LABORATORY

There are two general types of problem that are readily solved by electrophoretic techniques in the molecular laboratory. The first is the need to accurately assess the size of DNA fragments; the second is the need to identify both characterized and uncharacterized mutations or sequence variants.

5.1. SIZING OF DNA FRAGMENTS BY ELECTROPHORESIS The need to determine the size of DNA fragments is a common occurrence in the molecular laboratory. Examples include verification of the identity of cloned DNA fragments by restriction enzyme digestion, determination of the size of a band detected by Southern transfer, and verification of the size of a PCR product.

The size range of DNA fragments separable by any gel system is a function of the pore size of the matrix [this parameter is related to the viscosity of the matrix shown in Eq. (2) of Table 1]. The pore size is primarily a function of concentration in agarose gels and %T and %C in acrylamide gels. Even though the optimum separation ranges for different gels vary widely, they all have the same general profile. For all gels, there will be a region of optimal separation that is proportional to the log of the size of the DNA fragment. Similarly, there will be a point at which all DNA fragments are smaller than the effective pore size and will be not be retarded by the gel matrix and a point at which all DNA fragments are too large for any of the gel pores. At these two points, mobility is independent of molecular size, and all DNA fragments above or below these limits comigrate. In the laboratory, size determinations are typically made by comparison to a size marker that has been run on the same gel as the unknown sample. A variety of size markers containing DNA fragments of known length are commercially available. A standard curve is prepared by plotting the log of the migration distance of each band in the size marker vs its size in basepairs. The migration distance of the unknown band is plotted and the size of the fragment read off the graph.

After the discovery of trinucleotide repeat expansion as a mechanism for genetic disease, the molecular diagnostics laboratory had a need for high-resolution DNA fragment sizing. In this case, PCR products prepared with primers flanking the triplet repeat region of interest are electrophroresed on a high-resolution polyacrylamide gel or analyzed by CE. In the slab gel method, the high resolution is achieved by conducting the electrophoresis on a long (30–40 cm long), thin (0.4–0.6 mm thick) polyacrylamaide gel. In this technique, essentially identical to DNA sequencing methods, the PCR products are typically radiolabeled with 32P or 33P and detected by autoradiography. Figure 4 shows an example of a typical analysis of the Fragile X locus. In the CE method, one of the PCR primers is labeled with a fluorescent tag, and the sizing is done in reference to an internal standard, which consists of DNA fragments of known lengths that are labeled with a different fluorescent molecule. Figure 5 shows an example of a typical analysis of the spinocerebellar ataxia type I locus.

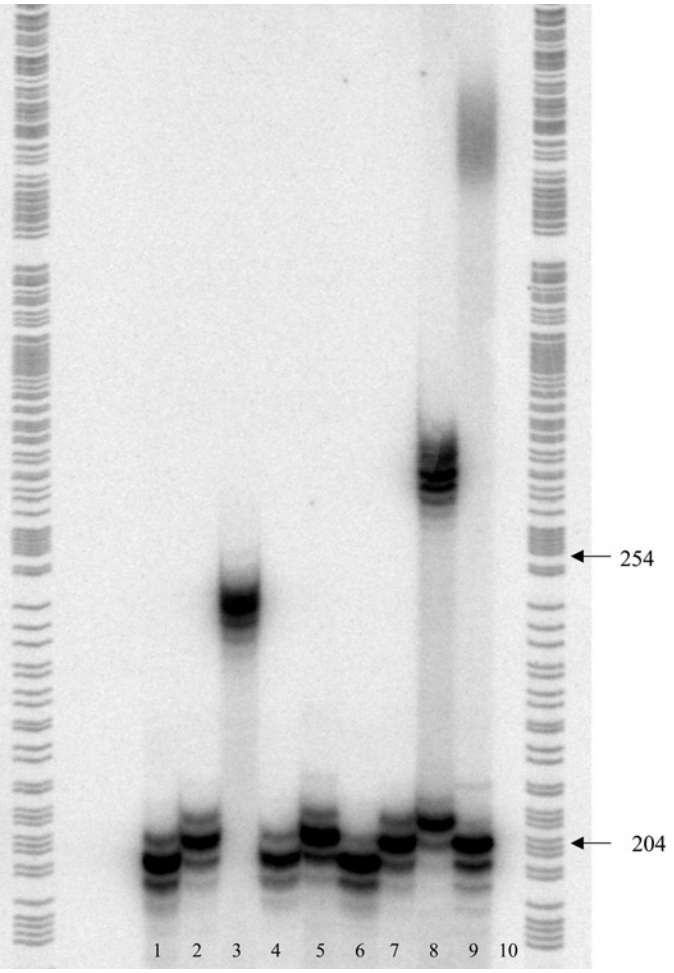


Fig. 4. Fragment-size analysis using slab gels and autoradiography: Fragile *X*. The autoradiograph shows PCR products using primers flanking the CCG repeat 5′ of the FMR1 gene. Sizes of patient alleles are determined in reference to the A lanes from a sequencing reaction using phage M13 as the template. The arrows indicate the position of the 204 and 254 base fragments of the sequencing ladder. Lanes 1, 4, and 6 show hemizygous males with amplicons of 200 bases, or 29 repeats. Lanes 2, 5, and 7 are from males with 30 repeats, Lane 3 shows a male with 43 repeats. Lane 8 shows a female heterozygous for a premutation allele of 52 repeats and a normal allele of 31 repeats. Lane 9 shows a control female sample with 30 and 87 repeats. Lane 10 is a no-DNA amplification control. (Courtesy of the Mayo Clinic Molecular Genetics Laboratory.)

5.2. SINGLE-STRANDED CONFORMATIONAL POLYMORPHISM ANALYSIS FOR DETECTION OF MUTATIONS AND SEQUENCE VARIANTS Single-stranded conformational polymorphism (SSCP) analysis is one of the most widely used mutation scanning systems for identification of unknown sequence variations. The reasons for its popularity are its high sensitivity to the presence of sequence variations and its technical simplicity. The extent of the popularity of SSCP can be appreciated from noting that a recent PubMed search of the term "SSCP" gave over 8000 hits. That many investigators worldwide have utilized this technique is appreciated by noting the number of PubMed citations over the years since 1989 (Fig. 6). The technique, developed in Hayashi's laboratory and first reported in 1989, involves PCR amplification of the region

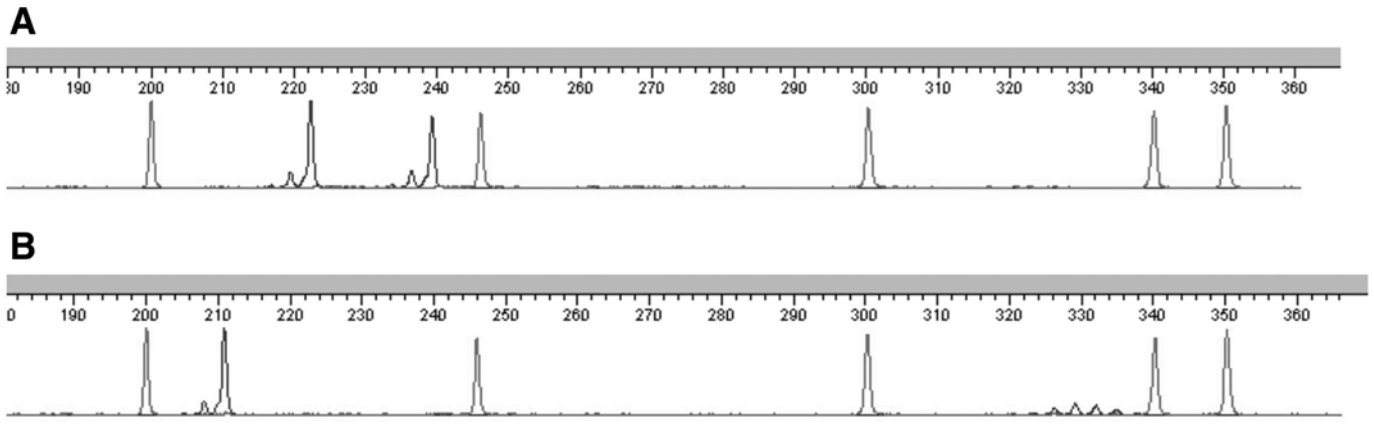


Fig. 5. Fragment-size analysis using CE and fluorescence detection: spinocerebellar ataxia type I. The electopherograms show PCR products using primers flanking the CAG repeat of the ataxin-1 gene. The forward PCR primer is labeled with 6-FAM; the resultant peaks are shown in blue. The PCR products are mixed with a ROX-labeled size marker, shown in red, and coinjected for capillary electrophoresis on an ABI 3100 (Applied Biosystems, Foster City, CA). **(A)** a patient with 2 normal alleles, 30 and 36 repeats (222 and 239 bases, respectively; **(B)** a patient with normal allele of 26 repeats (211 bases) and 1 expanded allele of 68 repeats (332 bases). This expansion is diagnostic of spinocerebellar ataxia type I. (Courtesy of the Mayo Clinic Molecular Genetics Laboratory.)

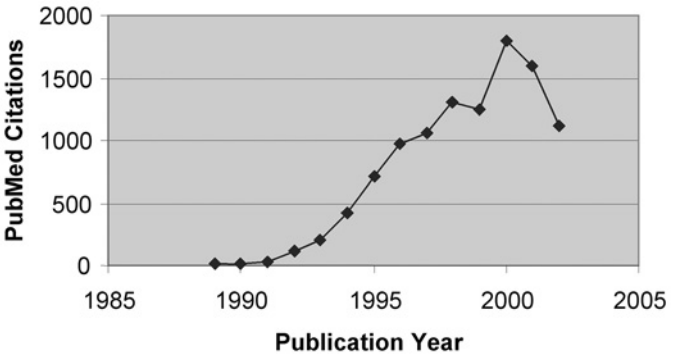


Fig. 6. The number of PubMed citations for SSCP by year since the original description in 1989.

of the gene to be studied, denaturation of the double-stranded PCR product by heat, and electrophoresis on a nondenaturing polyacrylamide gel *(15)*. During the electrophoresis, the single-stranded DNA fragments fold into a three-dimensional shape according to their primary sequence. The separation then becomes a function of the shape of the single-stranded molecules. If wild-type and mutant PCR products differ in their sequence, even by only a single nucleotide, they will likely adopt different three-dimensional structures and exhibit different electrophoretic mobilities (Fig. 7). Figure 8 shows SSCP analysis of exon 8 of the *p53* gene from three breast tumor cell lines *(16,17)*.

In order to prevent the two single strands from reannealing to form double-stranded DNA, the concentration of DNA in the loading buffer is kept very low. Thus, in order to visualize the bands, radioactive labeling of the DNA is typically required. However, as an alternative, silver staining seems to give adequate sensitivity and is increasing in popularity.

The issue of sensitivity of mutation scanning methods (i.e., the percentage of mutations detected) is a difficult one. Sensitivity is likely to be influenced by many factors, including, but not limited to; type of base substitution, length of the fragment examined, the local base sequences, the G+C content of the DNA fragment, and the location of the sequence variation relative to the ends of the fragment. For each of the mutation scanning methods, with the sole exception of denaturing gradient gel electrophoresis, there is no precise theory that can be used to predict whether a given method will detect a particular mutation. Thus, the only determinations of sensitivity of mutation scanning methods have been empirical. Typically, studies of sensitivity have used a set of previously characterized mutations found in the gene that the author is studying. Few of the above-listed variables are addressed in most studies, and no study to date has addressed them all.

The reported sensitivity of SSCP has ranged from 35% to near 100%. Sheffield and colleagues used a set of artificial mutants originally created by Myers et al. to study the effect of sequence variation on the promoter region of the mouse β-globin gene *(18,19)*. They demonstrated a pronounced effect of fragment length on sensitivity. At low fragment sizes (100–150 bp), SSCP had a sensitivity approaching 100%; however, when the length increased to 500 bp, the sensitivity dropped to approx 50%.

More recent studies by Highsmith and colleagues have attempted to examine a number of the parameters affecting the sensitivity of SSCP. In order to address all of these issues, this group prepared a set of plasmid constructs with inserts, derived from common cloning vectors, with G+C contents of 40%, 50%, and 60%. Then, using site-directed mutatgenesis, all four bases were introduced into the same position in each set of constructs. With a battery of PCR primers, the investigators could examine the effects of G+C content, fragment size, specific base change, and position of the mutation relative to the end of the fragment on sensitivity of SSCP. Using this "DNA Toolbox," this group found that the primary determinants of sensitivity were the G+C content of the PCR fragment and the temperature of the gel during electrophoresis *(20)*.

Running the SSCP gels under different conditions has been reported to increase sensitivity, as has running the gels on

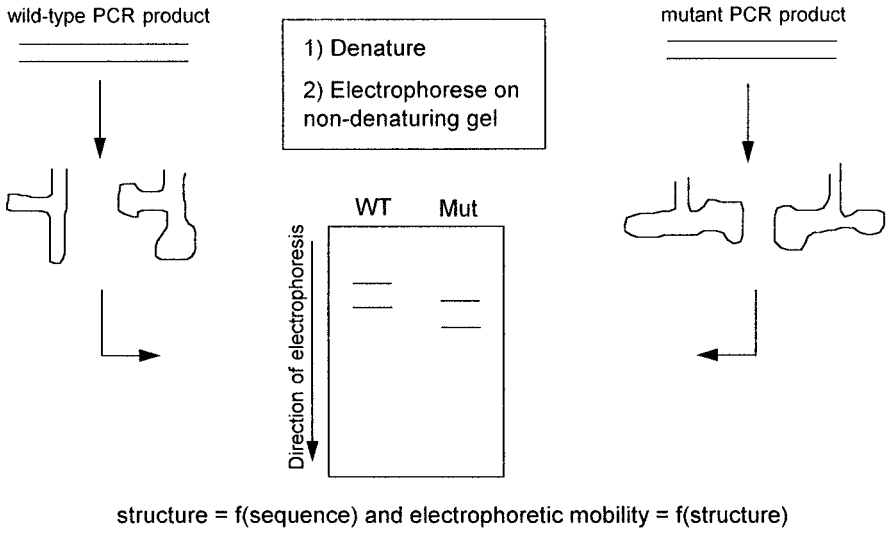


Fig. 7. Schematic representation of single-stranded conformational polymorphism analysis. In SSCP analysis, the electrophoretic mobilities of the single-stranded DNA species are a function of their three-dimensional conformation. This conformation is determined by the most thermodynamically favored intrastrand base-pairing. This, in turn, is directly determined by the primary sequence. Thus, if two DNA fragments differ in sequence, they will fold into different conformations and exhibit different electrophoretic mobilities.

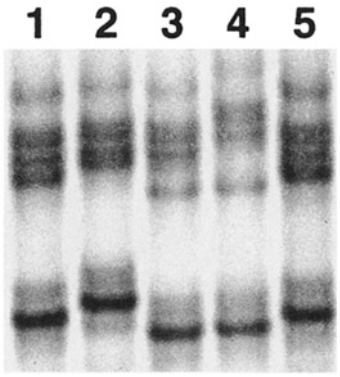


Fig. 8. Detection of *p53* sequence variants by sscp analysis. A 183-bp fragment of exon 8 of the *p53* gene was amplified from DNA extracted from four different breast tumor cell lines and human placenta (wild-type control). ^{33}P-labeled dATP was included in the PCR to label the products that were separated by electrophoresis through a 0.5X MDE gel at room temperature. Lane 1: Wild-type control; lane 2, MDA-MB-231 cells; lane 3, MDA-MB-468 cells; lane 4, Bt-474 cells; lane 5 SkBr3 cells. The cell lines in lanes 2–4 contain mutations in exon 8 resulting in aberrant migration of the PCR products. (Courtesy of Lori Terry and J. Carl Barrett, PH.D., Laboratory of Molecular Carcinogenesis, NIEHS.)

specialty matrices, such as MDE *(21)*. Currently, many large laboratories run SSCP gels under several electrophoretic conditions (e.g., at room temperature and at 4°C, and with gels containing 0% or 10% glycerol) The use of multiple-run conditions can raise the sensitivity of the technique to virtually 100% *(22)*.

Single-stranded conformational electrophoresis can be performed using CE as well as in the slab gel format. Although, at this writing, CE-SSCP has not been as thoroughly examined as SSCP using slab gels, several conclusions are beginning to emerge. As in slab gels, it appears that the fragment length is important in determining sensitivity of mutation detection, with longer fragments yielding decreased sensitivity *(23)*. The role of G + C content has not yet been examined in CE-SSCP. A wide variety of water-soluble, linear polymers have been used for CE-SSCP; however, there is yet to emerge a consensus as to which one is optimal. Because single-stranded DNA has more hydrophobic character than the double-stranded form, polymer hydrophobicity might prove to have a role in CE-SSCP separations. To date, only a single study has appeared examining the size of the polymer as a factor influencing mutation-detection sensitivity; that study *(24)* varied the molecular weight of linear polyacrylamide as the sieving matrix using the Molecular Dynamics MegaBase capillary array instrument (Sunnyvale, CA). The investigators compared linear polyacrylamides of molecular weight betweem 200 and 2400 kDa. They found that the lowest-molecular-weight solutions did not offer sufficient resolving power, whereas solutions of a 1000-kDA polymer gave good results. Increasing the polymer size to 2.4×10^6 Da offered no advantage; rather, very large polymers gave long analysis times without improved sensitivity and were difficult to load and flush from the capillary because of their very high viscosity.

Specific advantages of CE as a platform for SSCP include the high throughput and reproducibility. Investigators at the National Institute of Standards and Technology (NIST) have reported standard deviations of retention times in the range 0.05–0.5%. *(25,26)* By utilizing multiple fluors, multicapillary array instruments, and the multiwavelength fluorescence detectors, extremely high throughout can be achieved. Software for the interpretation (mutation calling) in an automated fashion will be key to implementing high-throughput mutation scans using SSCP. Such a piece of software has been recently introduced by Hayashi's group. The program QUISCA accomplishes tasks such as high-frequency noise filtering, baseline correction, peak calling, and peak quantification; it is freely available to academic users *(27)*.

Although the sensitivity of CE-SSCP can approach 100%, the specificity of the technique (indeed, of virtually all mutation

detection strategies) is much less studied. Investigators at NIST reported 3 false positives in an analysis of 15 samples, a disturbingly high percentage *(28)*. Others have reported false positives as a result of reannealing of PCR primers to the single-stranded DNA and the appearance of artifactual peaks *(29)*. Clearly, the rate of false positives in genetic screening is an issue that needs to be addressed.

5.3. HETERODUPLEX ANALYSIS IN MUTATION DETECTION Heteroduplex formation is required for many mutation scanning methods. Heteroduplex analysis (HA) refers to the technique based on the electrophoretic resolution of wild-type double-stranded DNA fragments from fragments of identical length and sequence, but having one basepair mismatch. The mismatch is formed when a wild-type DNA fragment is mixed with a mutant DNA fragment. When the mixture is heated to denature the double-stranded material and allowed to cool to reanneal the single-stranded molecules, four types of molecule result: (1) wild-type DNA, formed when the Watson strand of the wild-type reanneals with the Crick strand of the wild-type; (2) mutant DNA, formed when the Watson strand of the mutant reanneals with the Crick strand of the mutant; and (3) two heteroduplex species, formed when the Watson strand of the wild-type reanneals with the Crick strand of the mutant and when the Watson strand of the mutant reanneals with the Crick strand of the wildtype (Fig. 9).

Two types of heteroduplex exist, the structures of which have been predicted by Bhattacharyya and Lilley *(30)*. The first type is formed when the sequence difference between the two DNA fragments is one or more point mutations; the resulting heteroduplex is termed a "bubble"-type heteroduplex. The second type is formed when the sequence difference between the two fragments is a small insertion or deletion; the resulting heteroduplex is termed a "bulge"-type heteroduplex. Although "bulges" result in a large structural perturbation from the double-stranded homoduplex and are readily resolvable on polyacrylamide gels, the change in overall structure as a result of a "bubble" is much more subtle, and these heteroduplexes typically are not resolvable from the homoduplexes on agarose or polyacrylamide gels run under standard conditions (*see* Subheading 5.3.2).

5.3.1. Heteroduplex Analysis on MDE™ Gels In 1991, a communication from Bhattacharayya's laboratory reported the superior separation of heteroduplex species on Hydrolink D-5000 (AT Biochem) *(31)*. (*Note*: This gel matrix is no longer commercially available). As the proprietary formulation of D-5000 had not been developed for increased resolution of heteroduplexes, the results were a surprise to the scientists at AT Biochem who developed it. However, they quickly responded to the challenge of reformulating the material for optimum heteroduplex resolution. The resulting product, MDE™ (Mutation Detection Enhancement), has made HA analysis a viable technique for mutation detection. The development and properties of the Hydrolink series of gel matrixes have been described *(32)*. Figure 10 shows heteroduplex analysis of exon 10 of the *CFTR* gene. The samples with mutations and sequence variants are characterized by the presence of extra bands with aberrant migration.

As with SSCP, there is no theory that can be derived to predict which mutations will be detected by HA and which ones will not. In order to investigate the parameters affecting detection sensitivity of HA on MDE gels, Highsmith et al. conducted a study using the previously described DNA Toolbox. In contrast to SSCP, the G+C content of the DNA was not a significant parameter; rather, the detection sensitivity, over a range of PCR sizes from 100 to 600 bp, was most influenced by the specific mismatch. The detection rates and separation distances between heteroduplexes and homoduplexes generally followed the known thermodynamic instability of the mismatches, with G : G, C : C > A : G, T : C ≈ T : G, A : G > A : A, T : T *(33)*.

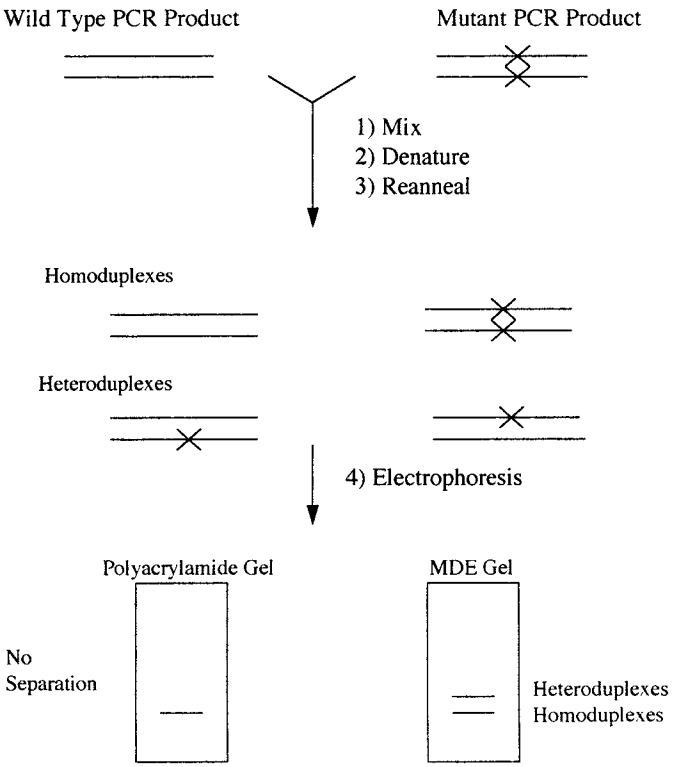


Fig. 9. Schematic representation of heteroduplex analysis. In HA mutations are detected by structural perterbations in the double-stranded DNA duplex because of the presence of one or more mismatches. Heteroduplexes as a result of single-base mismatches are not detected with high sensitivity on agarose or polyacrylamide, but they are typically resolved on specialty matricies such as MDE (Cambrex Bio Science Walkersville, Inc.).

5.3.2. Conformation-Sensitive Gel Electrophoresis Conformation-sensitive gel electrophoreis (CSGE) is a heteroduplex analysis technique described by Ganguly and colleagues in 1993 *(34)*. CSGE utilizes a polyacrylamide matrix with a novel crosslinker, bis-acrolylpiperazine (BAP). In addition, the CSGE matrix contains low concentrations of denaturing reagents. The concentration of these reagents is insufficient to cause complete separation of the DNA strands, but it is thought to further destabilize the conformational perturbations in the double-stranded DNA structure introduced by the presence of mismatched bases.

Like other electrophoretic mutation scanning methodologies, one of the primary determinants of sensitivity of CSGE is reputed to be DNA fragment length. In a recent update on CSGE, Ganguly reported a study in which the relative mobilities

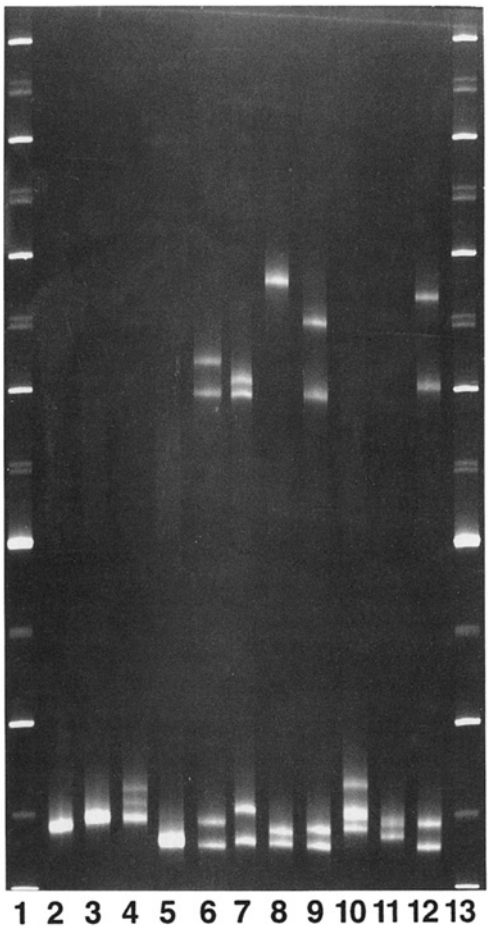


Fig. 10. Detection of mutations and sequence variants by heteroduplex analysis of the *CFTR* gene. Exon 10 of the *CFTR* gene was amplified by PCR for heteroduplex analysis. The gel is 1.5 mm thick and 40 cm long 1X MDE (Cambrex Bio Science Walkersville, Inc.) containing 15% urea. The gel was run for 21,000 V-h in 0.6X Tris–borate–EDTA buffer and stained with ethidium bromide. Lanes 1 and 13 are size markers (100 bp ladder), lane 2 is a MM homozygote at postion 470, lane 3 is a VV homozygote at position 470, lane 4 is a heterozygote for the M470V polymorphism, lane 5 is a homozygote for the ΔF508 mutation, lane 6 is a ΔI507/wild-type heterozygote, lane 7 is a compound heterozygote for ΔF508 and I506V, lane 8 is compound heterozygote for ΔF508 and F508C, lane 9 is a compound heterozygote for ΔF508 and Q493X, lane 10 is an I506V/wild-type heterozygote, lane 11 is a F508C/wild-type heterozygote, and lane 12 is a ΔF508/wild-type heterozygote.

of heteroduplexes vs homoduplexes in a set of constructs similar to the DNA Toolbox described earlier *(35)*. The PCR products ranged in size from 153 to 938 bp. A spike in resolution between heteroduplexes and homoduplexes was observed between 150 and 200 bp, with the resolution decreasing with increasing fragment size. However, sufficient resolution remained at 500 bp to be easily detectable. The greatly enhanced separations seen near 200 bp of fragment length were explained in terms of the model of DNA electrophoresis of Calladine et al. *(36)*, in which DNA is seen as a superhelix with inherent flexibility and a pitch of approx 200 bp. According to this interpretation, in longer DNA fragments, electrophoretic mobility is more influenced by bending than it is by local structural changes induced by mismatches.

One advantage of CSGE is the reproducibility of the patterns obtained with specific mutations. This can be of great help in the repetitive analyses of genes that have multiple benign polymorphisms as well as disease-causing mutations. Once the patterns of common polymorphisms are cataloged, there is no need to reflex samples exhibiting these patterns to sequencing. Figure 11 shows a typical CSGE gel interrogating three exons of the RET proto-oncogene in a screen for multiple endocrine neoplasia. A second advantage of CSGE over heteroduplex analysis on MDE gels is that the reagents are not proprietary. It would be of substantial interest to see a study comparing the two techniques for sensitivity and specificity of mutation detection. An advantage that CSGE and HA share over SSCP is that only one condition is needed and gels do not have to be run at subambient temperatures (i.e., in a cold room).

5.4. MUTATION DETECTION USING DENATURING GRADIENT GEL ELECTROPHORESIS Denaturing gradient gel electrophoresis (DGGE) was one of the first scanning methods used for the identification of mutations in DNA *(37)*. The method is based on the principle that the denaturation, or melting, of double-stranded DNA—by heat or denaturants such as hydroxide ion, urea, or formamide—does not occur in a single step; rather, DNA melts in domains. As the temperature rises or the denaturant concentration rises, the region or domain with the highest A + T content will melt first. If the temperature or the denaturant concentration is kept constant, the DNA structure composed of double-stranded DNA and a single region of single-stranded character will be stable. If the temperature is increased again, the region with the next highest A + T content will melt next. This melting by domain continues until the region with the highest G + C content is melted and the character of the DNA is completely single-stranded. As the identity of the melting domains is a function of the base sequence, a change in the base sequence (i.e., a mutation) will likely change the melting profile (Fig. 12). If the mutation does not alter the melting profile, it will not be detected. However, if a heteroduplex is formed and then subjected to melting analysis, a change in melting profile will almost certainly be seen. Thus, for maximum sensitivity, DNA heteroduplexes are formed between a control DNA fragment of known sequence and the test DNA.

The melting profile of heteroduplex DNA is observed by electrophoresis on a transverse denaturing gradient gel. In this system, a polyacrylamide gel is poured containing a gradient of denaturant, typically urea and formamide. After polymerization, the gel is rotated 90°, the DNA heteroduplex sample is applied to a single troughlike well, and the electrophoresis is performed. The DNA migrating through the region of the gel with the lowest denaturant concentration migrates as typical DNA; the DNA migrating through the denaturant concentration corresponding to the first melting domain will migrate with significantly lower mobility; the DNA migrating through the region corresponding to the next melting domain will migrate slower still. This stepfunction of decreasing mobility continues until the highest melting-point domain denatures, yielding rapidly moving single-stranded DNA. After visualization of the DNA with ethidium bromide or silver, a stair-step pattern is seen. In the case of a heteroduplex sample, the domain containing

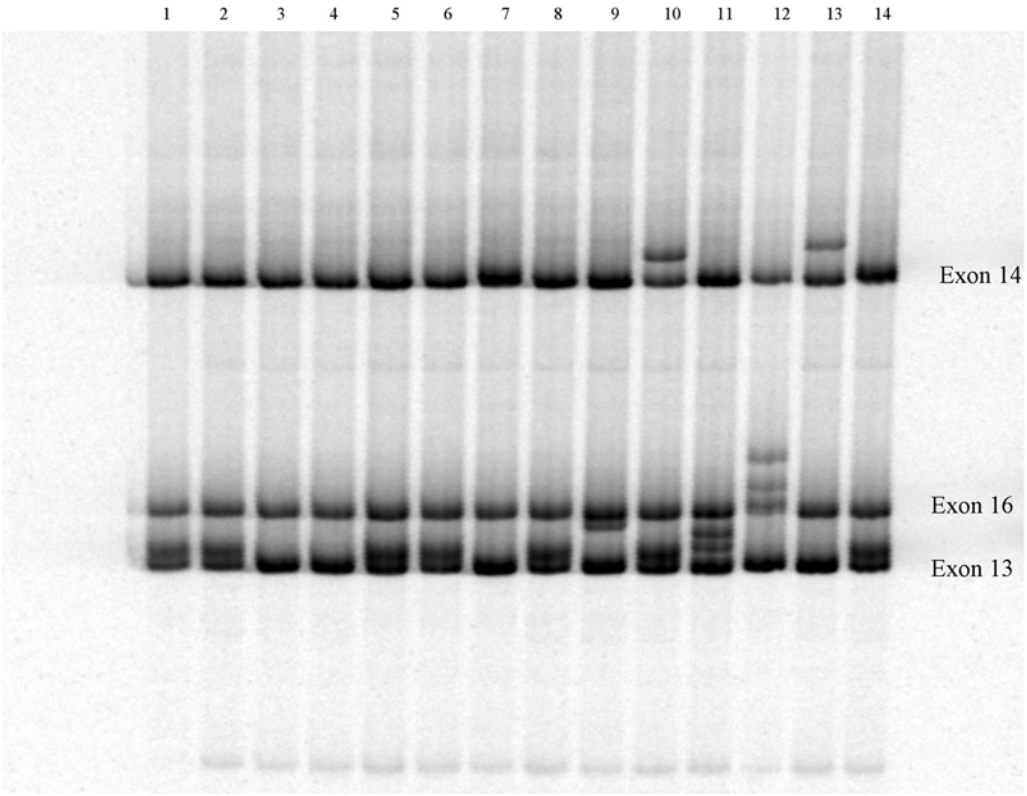


Fig. 11. Detection of mutations and sequence variants by CSGE analysis of the RET oncogene. Exons 13, 14, and 16 of the RET oncogene were amplified in a multiplex reaction containing ^{33}P-dATP. CSGE was performed as described by Ganguly *(35)*; bands were detected by autoradiography. Lanes 9–14 are controls for common polymorphisms. Lane 9 is a heterozygote for the E768D polymorphism. Lane 10 shows both L769L in exon 13 and V804L in exon 14. Lane 11 shows a sample bearing two polymorphisms in exon 13: E768D and L769L. Lane 12 shows M918T in exon 16. Lane 13 shows V804M in exon 14. Lane 14 shows L769L in exon 13 and S836S in exon 14. Lanes 1–8 are patient samples. Lanes 1, 2, 5, 6, and 8 are positive for the common L769L polymorphism. There are no pathogenic mutations identified on this gel.

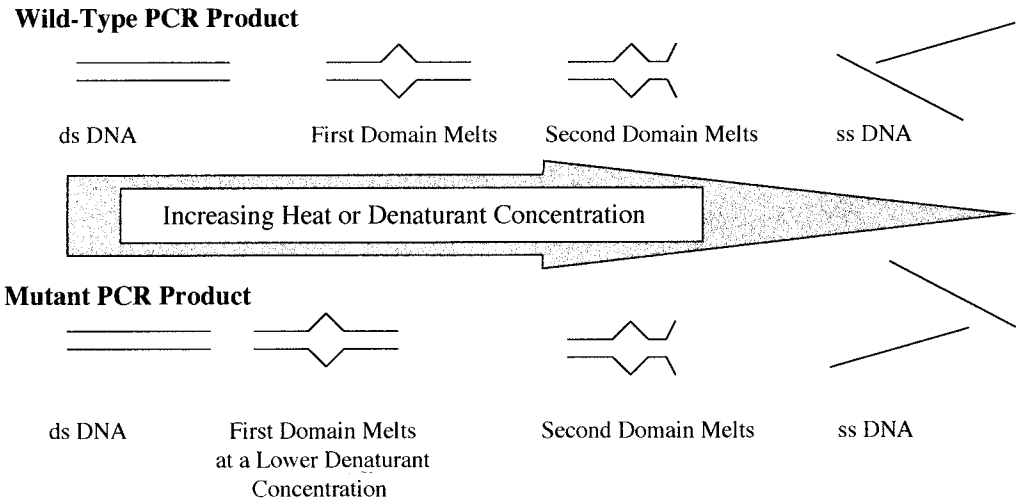


Fig. 12. DNA melts in domains. When treated with heat or chemical denaturants (urea or formamide), double-stranded DNA does not melt all at once, rather it melts in domains.

the mismatch will melt early, yielding a characteristic doublet pattern. (Fig. 13).

Clearly, pouring gradient gels and analyzing samples one at a time is labor-intensive and time-consuming. Fortunately, the melting profile of any DNA fragment can be modeled mathematically. Computer programs are commercially available that calculate the melting profile for PCR-amplified DNA *(38)*. Using this tool, gradients can be optimized for each piece of DNA. Using optimized gradients, gels can be run in a more conventional manner (i.e., with the electrophoresis driving the

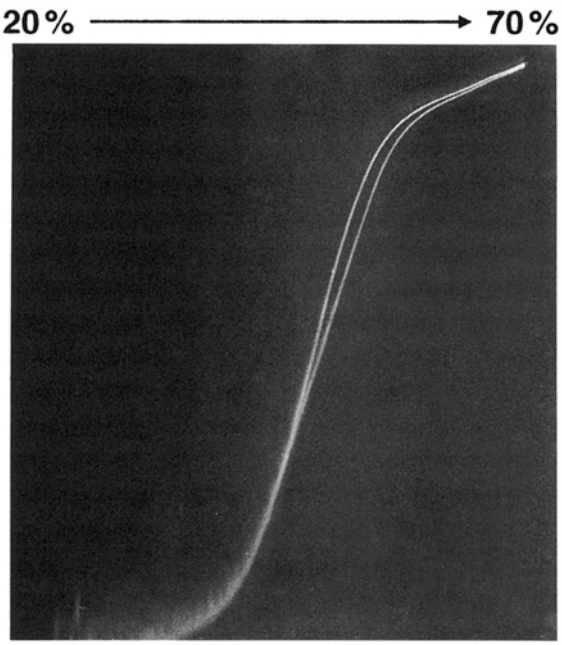


Fig. 13. A transverse DGGE gel of exon 10 of the *CFTR* gene from an individual heterozygous for the ΔF508 mutation. The direction of electrophoresis is top to bottom and the denaturant concentration increases from left to right. The heteroduplex product melts at a lower temperature than the first domain of the homoduplexes. Thus, the mobility transition occurs further to the left (lower denaturant concentration) than for the homoduplexes, giving the characteristic "cats-eye" pattern, indicative of the presence of a mixture of sequences.

DNA into ever higher denaturant concentrations). In this format, a sample bearing a low melting-point domain, such as a heteroduplex, will exhibit a band of slower mobility.

In theory, mutations in the highest melting-point domain will not be detected by DGGE because it is difficult to electrophoretically determine the exact point at which the transition from the slow migrating species (with the last melting-point domain intact) to the rapidly migrating single-stranded species occurs. In order to eliminate this caveat to the technique, a "GC clamp" is typically added to the amplified product by preparing one of the PCR primers with a 5′ tail of 30–50 nucleotides of 100% G+C content. Thus, the artificial "GC clamp" becomes the highest melting-point domain, and mutations in all of the original domains can be detected *(39)*. Although the sensitivity of DGGE for the detection of unknown mutations approaches 100%; the popularity of the method seems to be decreasing because of the expense of PCR primers that are 70–80 nucleotides in length, the difficulty in amplification with primers of such high-melting temperatures, and the difficulties associated with the need to freshly prepare reproducible gradient gels. However, an experimentally more facile system, temperature sweep or temporal thermal gradient gel electrophoresis (TTGE), was introduced by Yoshino et al. in 1991 *(48)*. This system, which relies on the same principle as DDGE, is far more straightforward to carry out, as it does not involve pouring gradient gels. In a TTGE experiment, the denaturing gradient is provided as a smooth increase in temperature during the electrophoretic run. Wong et al. recently reported the use of TTGE in a comprehensive scan of the entire mitochondrial genome from 179 patients for deleterious mutations and polymorphisms. *(49)*. A system for TTGE, Dcode™, is commercially available form Bio-Rad (Hercules, CA).

5.5. MUTATION DETECTION BY CHEMICAL CLEAVAGE OF MISMATCHED NUCLEOTIDES The chemical cleavage of mismatches technique, described by Cotton in 1988, takes advantage of the differential reactivity of perfectly paired and mismatched bases to chemical modifying reagents *(50)*. In heteroduplex species in which a thymine nucleotide is mismatched, the T residue is hypersusceptible to chemical modification by osmium tetroxide (OsO_4), a commonly used shadowing reagent for electron microscopy. Similarly, mismatched cytosine nucleotides are hypersusceptible to attack by hydroxlamine ($HONH_2$). DNA strands containing either a modified T or C nucleotide are then cleaved with piperidine ($C_5H_{11}N$). In practice, the DNA to be screened for mutations is amplified and then mixed with a 5- to 10-fold molar excess of wild-type amplicon. This control DNA, referred to as the probe, is typically labeled on one strand with ^{32}P. After mixing, melting, and reannealing, the resultant heteroduplexes are divided into two aliquots. One aliquot is treated with OsO_4 and the other with $HONH_4$. After treatment with these reagents, the samples are treated with piperidine and separated by electrophoresis on a sequencing-type polyacrylamide gel. If a mutation is present, it will be detected as an extra band after autoradiography. If the sample is tested twice (once with each strand of the probe DNA labeled), virtually 100% of all mutations will be detected. Furthermore, the exact position of the mutation can be defined by sizing the cleavage product. Sensitivity of CCM for the detection of mutations is very high (95–100%), the toxicity of the reagents, the large number of steps and manipulations, and the high background seen with many templates has limited the number of laboratories that have used this technique (reviewed in refs. *40* and *41*).

5.6. MUTATION DETECTION BY RIBONUCLEASE CLEAVAGE OF MISMATCHED RNA : DNA DUPLEXES Mutation detection based on ribonuclease cleavage was developed when it was recognized that ribonucleases could cleave single-stranded RNA and that it was possible to synthesize radioactive RNA probes *(42)*. RNA probes, or riboprobes, are synthesized using wild-type genomic DNA as a template with ^{32}P incorporated as a label. The probe is hybridized to denatured target DNA to produce RNA : DNA hybrids. When there is a mismatch between the wild-type RNA probe and the DNA because of a mutation, the base or bases that have not annealed to the DNA are cleaved by RNAase A. The products of the digestion are denatured and separated by gel electrophoresis. A mutation is indicated by the presence of cleavage fragments of lower molecular weight than the full-length probe. The size of the cleavage products are used to determine the location of the specific mutation. Although this technique has been used to detect mutations in the hypoxanthine phosphoribosyltransferase, type I collagen, and K-*ras* genes, it has not been widely employed partly because of the inability of RNAase A to completely cleave all mismatches *(43–45)*. Single-stranded RNA beause of mismatches involving the purines adenine and guanine are not efficiently recognized by RNAase A. However,

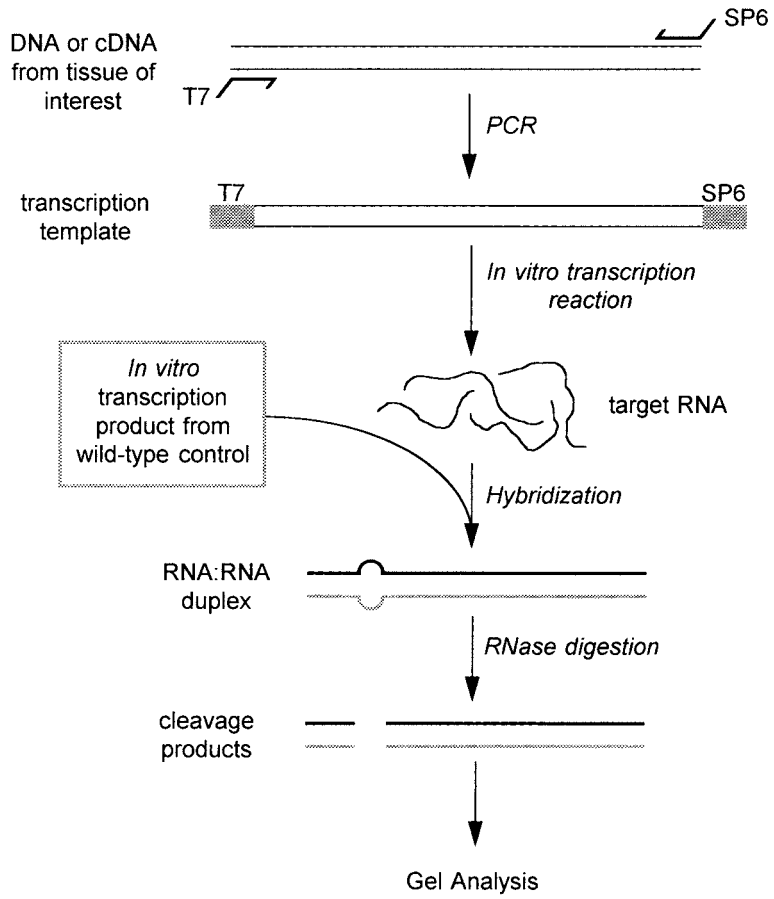


Fig. 14. Schematic representation of a nonisotopic RNase cleavage assay. The DNA or cDNA is amplified using primers with 5′ T7 or SP6 promoter sequence to generate a template for in vitro transcription. The template, or target RNA, is hybridized with wild-type RNA to produce RNA : RNA duplexes. When a mismatch is present because of a mutation in the target RNA, the duplexes are cleaved by RNAase. The digested products are separated by gel electrophoresis and visualized by staining with ethidium bromide. (Courtesy of Marianna M. Goldrick, Ph.D., Ambion, Inc., Austin, TX.)

mismatches involving the pyrimidines cytosine and uracil or larger areas of single-stranded RNA (because of two mismatched sites in close proximity, a deletion or an insertion) are effectively cleaved. Modification and improvement of this method was made by incorporating PCR amplification of the target sequence and use of RNAase I *(46)*. The starting template is either mRNA or genomic DNA. It is advantageous to use mRNA as a template because there are no intron sequences present. RNAase I recognizes all 4 bases when they are present at the site of a mismatch. However, like the original RNAase cleavage mismatch protocols, incomplete digestion of the hybrid molecules makes it difficult to distinguish between homozygotes and heterozygotes.

A nonisotopic RNase cleavage assay for mutation detection has been developed by Ambion, Inc., Austin, TX *(47)*. Beginning with DNA or cDNA, the sequence to be screened is amplified using a forward primer with a T7 bacterial promoter sequence and a reverse primer with a SP6 bacterial promoter sequence (Fig.14). Target segments up to 1.0 kb long can be amplified and screened in one reaction. The added bacterial promoter sequences allow the PCR products to be transcribed in an in vitro system to produce large quantities of target RNA. The target RNA is hybridized with wild-type RNA to form RNA : RNA duplexes. Because both template strands are transcribed, reciprocal mismatches (i.e., A-C and G-U) are created when each strand hybridizes to the wild-type RNA. This increases the likelihood that a mismatch will be cleaved because all sites are not cleaved with equal efficiency.

The hybrids are treated with RNAase, and any unpaired mismatched residues accessible to the enzyme are cleaved. The cleavage products are stained with ethidium bromide, separated by gel electrophoresis, and compared to the wild-type homoduplex, which was also treated with enzyme. The wild-type homoduplex should be a single band of the highest molecular weight because it is resistant to cleavage. Although it is not possible to determine if the cleavage is at the 5′ or 3′ end of the target segment without rescreening, the size of the cleavage product does give a good estimation of the position of the mutation. A significant advantage of this method is the visualization of the cleavage products without radioactive probes. As shown in Fig. 15, this method can be used to screen for germline mutations in the breast and ovarian cancer susceptibility gene *BRCA1*. In addition to the detection of mutations, a nonisotopic Rnase cleavage assay (NIRCA) can also identify the genotype of a sample by hybridization of the sample to its own RNA transcripts. If a sample is heterozygous for a mutation,

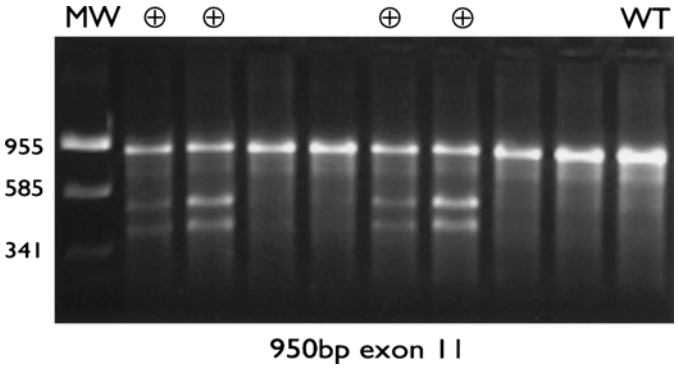


Fig. 15. Detection of mutations in the BRCA1 gene using a nonisotopic RNase cleavage assay. Target regions of the *BRCA1* gene were amplified by nested PCR from genomic DNA from at-risk relatives of a familial breast cancer patient and from a normal control. The primers had promoter (T7 and SP6) sequences on the 5′ ends. Crude PCR products (2 μL) were transcribed with T7 and SP6 RNA polymerase. Complementary normal and test transcripts were mixed, heated briefly and cooled to make double-stranded RNA targets. Targets were treated with RNase for 45 min at 37°C, separated by electrophoresis through a 2% agarose gel, and then stained with ethidum bromide. Samples in lanes marked ⊕ were scored as positive for a putative mutation. The WT lane shows wild-type control sample. DNA size markers are indicated at the left margin. (Courtesy of Marianna M. Goldrick, PH.D., Ambion, Inc. Austin TX.)

self-hybridization results in mismatched hybrids, which are cleaved. However, if a sample is homozygous, self-hybridization results in completely matched duplexes that are resistant to cleavage. The use of RNase cleavage methods for the detection of mutations has recently been reviewed by Goldrick *(51)*.

6. TECHNIQUES FOR DETECTION OF KNOWN MUTATIONS AND SEQUENCE VARIANTS

Restriction endonucleases have proven to be extremely useful in the analysis and characterization of PCR products. Digestion of PCR products with endonucleases can be used to confirm the amplification of desired sequences when the size of the fragments can be predicted from known restriction sites. In addition, it can be used to identify sequence variants and provide linkage information for pedigree analysis when a mutation has not been identified or before a gene is cloned *(52,53)*.

Restriction endonucleases protect bacteria from invasion by foreign DNA by recognizing and cleaving specific sequences in double-stranded DNA *(54)*. The bacteria's own DNA is protected from digestion through methylation or modification of the restriction sites; thus, they are not recognized by the enzyme. There are three classes of endonuclease with different cofactor requirements and different DNA recognition abilities. The class II enzymes are most commonly used in molecular biology applications. They require only the presence of Mg^2+, recognize DNA sequences approx 4–8 bases long and cleave at, or near, the recognition site. Many of these sites are palindromic, and when cleaved, they result in "blunt" or "sticky" ends. These new ends are extremely advantageous for ligation of the fragment into vectors for further manipulation. The appropriate digestion conditions and buffer for each enzyme are usually supplied by the manufacturer. Because most enzymes are active at 37°C, several can be combined in one reaction if a single product is to be cut at several sites or if a multiplex reaction contains several different products with different restriction sites. The resulting fragments are separated on an agarose or acrylamide gel, depending on the required resolution and stained with ethidium bromide for visualization.

There are numerous restriction sites throughout any region of DNA. Some of these sites are polymorphic in that on a given allele, the site might be present or absent. This could be part of normal variation and not result in any disease. The presence or absence of the site affects whether the DNA fragment is cleaved by an endonuclease. If the polymorphism is closely linked to a disease locus, in some families it could be used as a marker to follow inheritance of the mutant (disease-producing) allele. Prior to the development of direct detection methods to identify cystic fibrosis mutations, analysis of linked restriction fragment length polymorphisms (RFLPs) was used for prenatal diagnosis *(53)* and is still useful when both mutations in the parents have not been determined. This type of analysis requires that an affected individual and both parents be available for testing to ascertain which parental alleles carry the mutant genes. DNA obtained from chorionic villus sampling or amniocentesis is the template for PCR amplification of the region encompassing the polymorphism. Based on the restriction fragment patterns resulting from digestion of the PCR products, the genotype of the fetus is determined. Linkage marker analysis for prenatal diagnosis is not 100% accurate because of the possibility of recombination between the alleles during meiosis. Therefore, it is desirable to analyze several markers to increase the certainty of the diagnosis.

Once a disease gene is identified and mutations have been detected and characterized, RFLP analysis might be used to screen samples for the same mutations if they create/destroy a restriction site. The sequence flanking the mutation is amplified and digested with the appropriate enzyme, and the fragments are resolved by gel electrophoresis. For example, the cystic fibrosis mutation 2789 + 5G > A creates a *Ssp*I cutting site. When a 305-bp region of the gene that encompasses this mutation is amplified and digested, three fragments result. (One additional fragment results from a constitutive site present in both alleles.) The wild-type allele is cut only once at the constitutive site to produce two fragments (Fig. 16) *(55)*. Mutations and sequence variants that alter restriction sites have been identified in virtually all genes studied.

Unfortunately, a sequence variant does not always change a restriction site. Therefore, in order to preserve the simplicity of mutation detection by PCR/restriction digestion,

PCR-mediated site-directed mutagenesis can be applied *(56)*. This technique creates or destroys restriction sites in the PCR product by introduction of a base substitution near the mutation by modifying the primers. This allows the detection of point mutations as well as small insertions and deletions that cannot be resolved through gel fractionation of the PCR products. A polymerase must be used that does not have exonuclease activity, or the mismatched primer will be corrected. The mismatched base can be several bases from the 3′ end of the primer to stabilize the primer–template hybrid without decreasing the efficiency of the polymerase or the specificity of the amplification. Several cystic fibrosis mutations can be detected

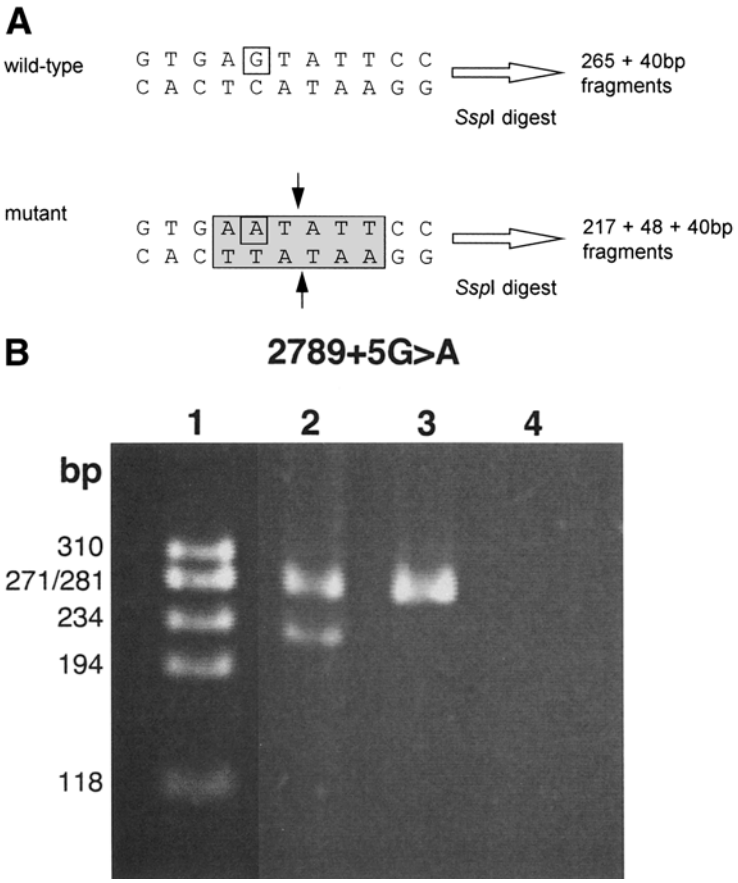


Fig. 16. Restriction fragment length polymorphism analysis of the cystic fibrosis mutation 2789+5G>A. **(A)** The wild-type PCR product is digested with *Ssp*I to two fragments resulting from the presence of a constitutive restriction site not associated with the mutation. When the G (surrounded by a small box) is mutated to an A, an additional *Ssp*I site is created to yield three fragments. The *Ssp*I site is indicated by the shaded box and the cutting site is indicated by the arrows. **(B)** A 305-bp region of the *CFTR* gene that flanks the 2789+5G>A mutation was amplified and digested with *Ssp*I. The PCR products were separated by electrophoresis through a 4% agarose gel and then stained with ethidum bromide. Lane 1: ϕX174 DNA size markers; lane 2, individual heterozygous for the 2789+5G>A mutation (265- and 217- bp fragments); lane 3, normal individual (265bp fragment); lane 4, water blank. The 40- and 48-bp fragments migrate quickly through the gel and are not visible. (Courtesy of Michelle L. Blalock, Molecular Genetics Laboratory, University of North Carolina Hospitals.)

by introduced restriction sites with mismatches 1, 2 or 3 bases from the 3′ end *(57,58)*. An example of an assay for the G542X mutation is shown in Fig.17.

Failure of the endonuclease to digest a PCR product can lead to misleading results in RFLP or PCR-mediated site-directed mutagenesis assays. Control samples, homozygous and heterozygous for both alleles, should be amplified and digested at the same time as the unknown. Additional endogenous or engineered restriction sites within the PCR product can act as an internal control for complete digestion. In some instances, such as the G542X and 2789 + 5G > A cystic fibrosis mutations, a constitutive site within the PCR product is valuable as an internal control for digestion *(57)*. In cases in which a constitutive site is not found on the amplicon being investigated, coamplification of a sequence containing the appropriate restriction site in an irrelevant gene for use as a restriction digest control is highly recommended.

The decision to introduce a restriction site into either allele is usually based on the location of the mismatch from the 3′ end of the primer and the commercial availability of the enzyme. However, the assay is more specific for a particular mutation when the primers are designed to introduce a cutting site into the mutant allele rather than ablate an existing one *(57,59)*. The reason for this is that if the restriction site is removed in the mutant allele, a rare polymorphism in the restriction site can also prevent the digestion of the PCR product, giving a false-positive result.

7. DNA SEQUENCING

DNA sequencing is considered the "gold standard" for mutation detection. However, before proceeding with a description of methods and technologies, it is important to point out that sequencing is not perfect, and it certainly is not magic. Like any analytical method, one can encounter false positives and false negatives. However, when properly done and properly interpreted, sequencing a DNA fragment will almost always reveal a high percentage, near 100%, of the sequence variations present.

Conceptually, three steps are required in order to obtain the base sequence of any DNA fragment. First, starting from some defined point, (e.g., the 5′ end of a PCR product or an annealed primer), one must separate the DNA into four reactions and

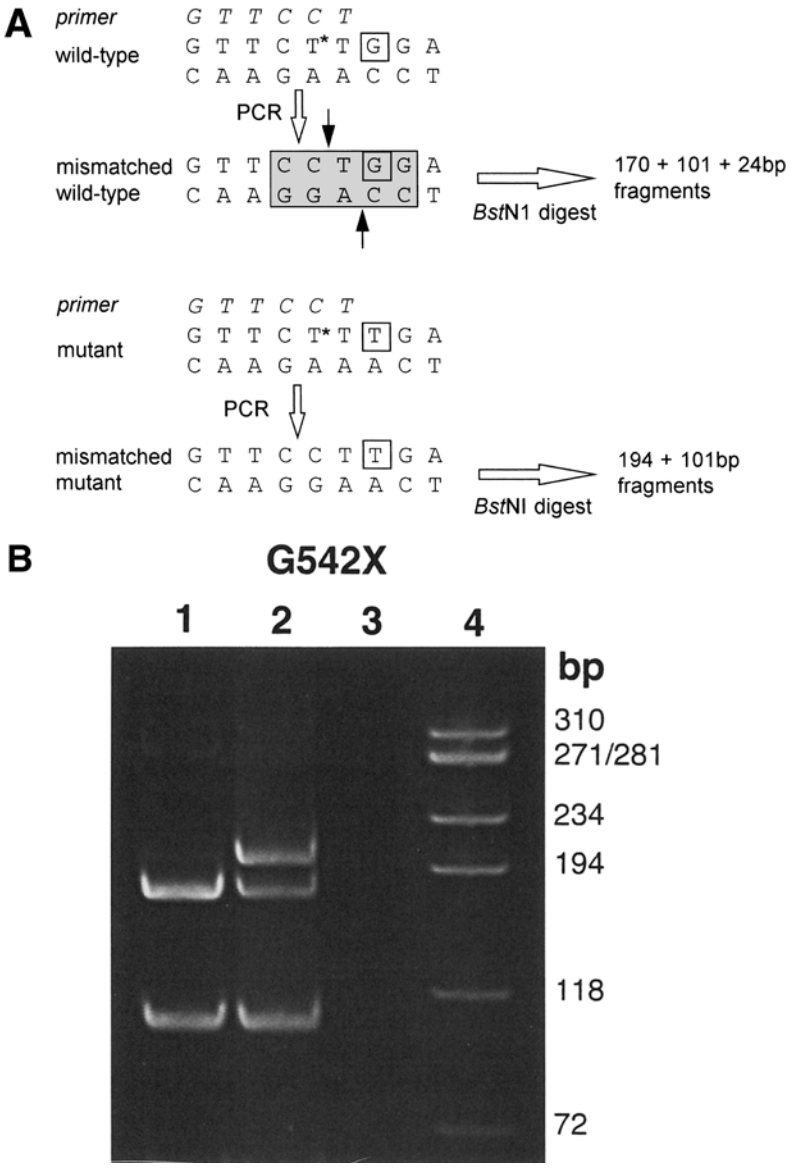


Fig. 17. PCR-mediated site-directed mutagenesis analysis of the cystic fibrosis mutation G542X. **(A)** A *Bst*NI site is introduced into the wild-type PCR product by changing the T (indicated by an asterisk) to a C during PCR amplification using mismatched primers. The new *Bst*NI site is indicated by the shaded box and the cutting site is indicated by the arrows. *Bst*NI digestion produces three fragments as a result of the presence of the introduced restriction site and a constitutive restriction site not associated with the mutation. When the wild-type G (surrounded by a small box) is mutated to a T, a *Bst*NI site is not created by the mismatched primers. The PCR product with the mutation is cut only once at the constitutive restriction site to produce two fragments. **(B)** A 295-bp region of the *CFTR* gene that flanks the G542X mutation was amplified and digested with *Bst*NI. The PCR products were separated by electrophoresis through a 10% acrylamide gel and then stained with ethidum bromide. Lane 1, normal individual (170- and 101-bp fragments); lane 2, individual heterozygous for the G542X mutation (195-, 170-, and 101-bp fragments); lane 3, water blank; lane 4, ϕX174 DNA size markers. The 24-bp fragment migrates quickly through the gel and is not visible. (Courtesy of William G. Learning, Molecular Genetics Laboratory, University of North Carolina Hospitals.)

specifically cleave it at the each of the four nucleotide bases; that is, in the A reaction, the DNA must be broken into all of the possible oligonucleotides ending in an A, the G reaction must produce all of the possible oligonucleotides that end in a G, and so on. Second, these oligonucleotides must be labeled in some manner to enable detection. Third, the labeled fragments must be separated, at single-base resolution, and detected.

In 1977, two articles appeared that demonstrated how to carry out these three steps and achieve the objective of reading a DNA sequence. The two methods accomplished the goals of creating pools of oligonucleotides, each pool consisting of DNA fragments ending in one of the four bases, in quite different ways. Maxam and Gilbert used chemicals to fragment DNA at specific bases *(60)*. Sanger and colleagues used the DNA to be sequenced as a template and synthesized four pools of new oligonucleotides, each one terminating at each of the four bases by incorporation of a chain-terminating dideoxynucleotide triphosphate *(61)*. These two publications are considered to be among the most significant milestones in molecular biology. In

1980, Gilbert and Sanger received the Nobel Prize in Chemistry "for their contributions concerning the determination of base sequences in nucleic acids." Although the Maxam–Gilbert chemical method has applications in sequencing small DNA fragments, such as oligonucleotides, currently, all, or almost all, DNA sequencing is done using the enzymatic Sanger method. Therefore, the rest of this discussion will focus exclusively on the enzymatic Sanger method.

The Sanger sequencing reaction is a multistep process. As most often practiced in the modern diagnostic laboratory, the first step is the PCR amplification of the target of interest. Clearly, this is a critical step, and a poorly designed PCR strategy with nonspecific amplification or poor yields will result in uninterpretable sequence. The next step involves the removal of excess deoxynucleotide triphosphates (dNTPs) and PCR primers. Next, the PCR product is denatured and an oligonucleotide (the sequencing primer) is added and annealed 5′ to the region to be sequenced. A DNA polymerase (typically a recombinant thermostable enzyme) and a mixture of dNTPs and dideoxynucleotide triphosphates (ddNTPs) is then added (and subjected to thermal cycling if a thermostable enzyme is used). The DNA polymerase will extend the annealed primer in the 5′ to 3′ direction, making a new strand of DNA that is complementary to the PCR product template. Because the ddNTPs retain a 5′ hydroxyl group, they can be incorporated into the growing complementary strand of DNA. However, as they lack the 3′ hydroxyl, they cannot be further extended by DNA polymerase. Thus, for example, when a ddATP is incorporated, the chain terminates at that A position, complementary to the corresponding T in the template. The rate at which ddNTPs are incorporated in the growing strand is dependent on both the ratio of dNTPs to ddNTPs in the reaction and the efficiency with which the ddNTPs are recognized by the polymerase. Most applications in the clinical laboratory utilize dye-terminator chemistry. In this approach, the detectable label, in this case a fluorescent tag, is linked to the ddNTPs. Each of the four ddNTPs is labeled with one of four fluorescent molecules. Thus, the fragment terminated by incorporation of a ddATP will be labled with one color, a fragment terminating in a G will be labeled with another color, and so on. The pool of DNA fragments is then subjected to high-resolution electrophoresis, either slab gel or CE, and the presence of each of the four fluors is scored as the DNA fragments migrate past a fixed, multiwavelength fluorescence detector. The output resembles a chromatogram and is read from the shortest fragment to the longer fragments, with a blue (for example) peak being read as an A, a red peak being read as a C, and so forth. A schematic of the principle modifications of the Sanger sequencing chemistry is shown in Fig. 18.

There have been enormous technical advances in DNA sequencing technology since the first reports. In the early days of sequencing, the ability to read 100 bases after several days of work was the state of the art. Today, automated sequence analyzers capable of turning out sequence at the rate of up to 100,000 to 200,000 bases per day are available and, although expensive, are within the reach of large clinical laboratories.

7.1. ENZYMES USED FOR DNA SEQUENCING One area in which major advances have occurred is the area of enzyme technology *(62)*. The original enzyme used by Sanger et al. was the Klenow fragment of *Escherichia coli* DNA polymerase I. The Klenow fragment retains the 5′ to 3′ DNA polymerase activity of the polymerase holoenzyme, but lacks much of the 3′ to 5′ exonuclease, or proofreading activity. Enzymes that retain the 3′ to 5′ exonuclease activity are not useful in DNA-sequencing applications for several reasons. First, it is difficult to achieve maximum labeling because the exonuclease activity can degrade the sequencing primer. Second, the exonuclease can remove the labeled dye terminator, allowing the new strand to continue growing by the addition of dNTPs and leaving gaps in the sequence. Third, the polymerase can pause at certain sequences and cycle back and forth between the polymerase and exonuclease activities. This can lead to marked variability in peak intensity as ddNTPs are given many more chances to incorporate at a pause site *(63)*.

The Klenow fragment was used for sequencing for several years and is still used on occasion. However, its low processivity and relatively low elongation rate leads to higher backgrounds and lower signals than some other enzymes. Sequenase™ is the trade name of T7 DNA polymerases developed by Tabor and Richardson. Sequenase version 1 was a chemically modified enzyme that retained the high processivity and eleongation rate of the native polymerase while possessing approx 1% of the original 3′ to 5′exonuclease activity *(64)*. Sequenase version 2 is a genetically modified enzyme that retains even less exonuclease activity *(65)*. The addition of manganese ions to the sequencing reaction virtually eliminates the incorporation efficiency differences between dNTPs and ddNTPs, resulting in extremely uniform bands *(66)*. Sequenase is marketed by US Biochemicals (Cleveland, OH). Thermal cycling and the use of thermostable DNA polymerases has certain advantages for DNA sequencing. In contrast to the PCR reaction, in which exponential amplification is achieved by the use of two primers, DNA sequencing uses a single primer. Thus, thermal cycling results in linear amplification of labeled sequencing products. This 20- to 30-fold amplification is sufficient to greatly reduce the amount of template required for each reaction. In addition, because fluorescent-detection methods are not as sensitive as radioactive detection, this linear amplification is particularly useful for fluorescent sequencing. *Taq* polymerase, the most commonly used enzyme for PCR, does not have significant 3′ to 5′ exonuclease activity. It does, however, have 5′to 3′ exonuclease activity. When used for sequencing, enzymes with this activity can degrade the labeled sequencing products from the 5′ end, resulting in undesirable length heterogeneity of products that retain the 3′ fluorescent label. Tabor and Richardson developed two genetically modified *Taq* polymerases in 1995 *(67)*. One of them, marketed under the name Amplitaq FS® by Appied Biosystems (Foster City, CA), has a single point mutation that virtually eliminates the 5′ to 3′ exonuclease activity. The other, marketed under the name Thermo-Sequenase® by Amersham Biosciences (Piscataway, NJ) and US Biochemicals (Cleveland, OH), contains another point mutation that greatly reduces the discrimination against ddNTPs, resulting in more even band intensities.

7.2. LABELING OF DDNTP TERMINATED FRAGMENTS Another area in which great technical advances have occurred

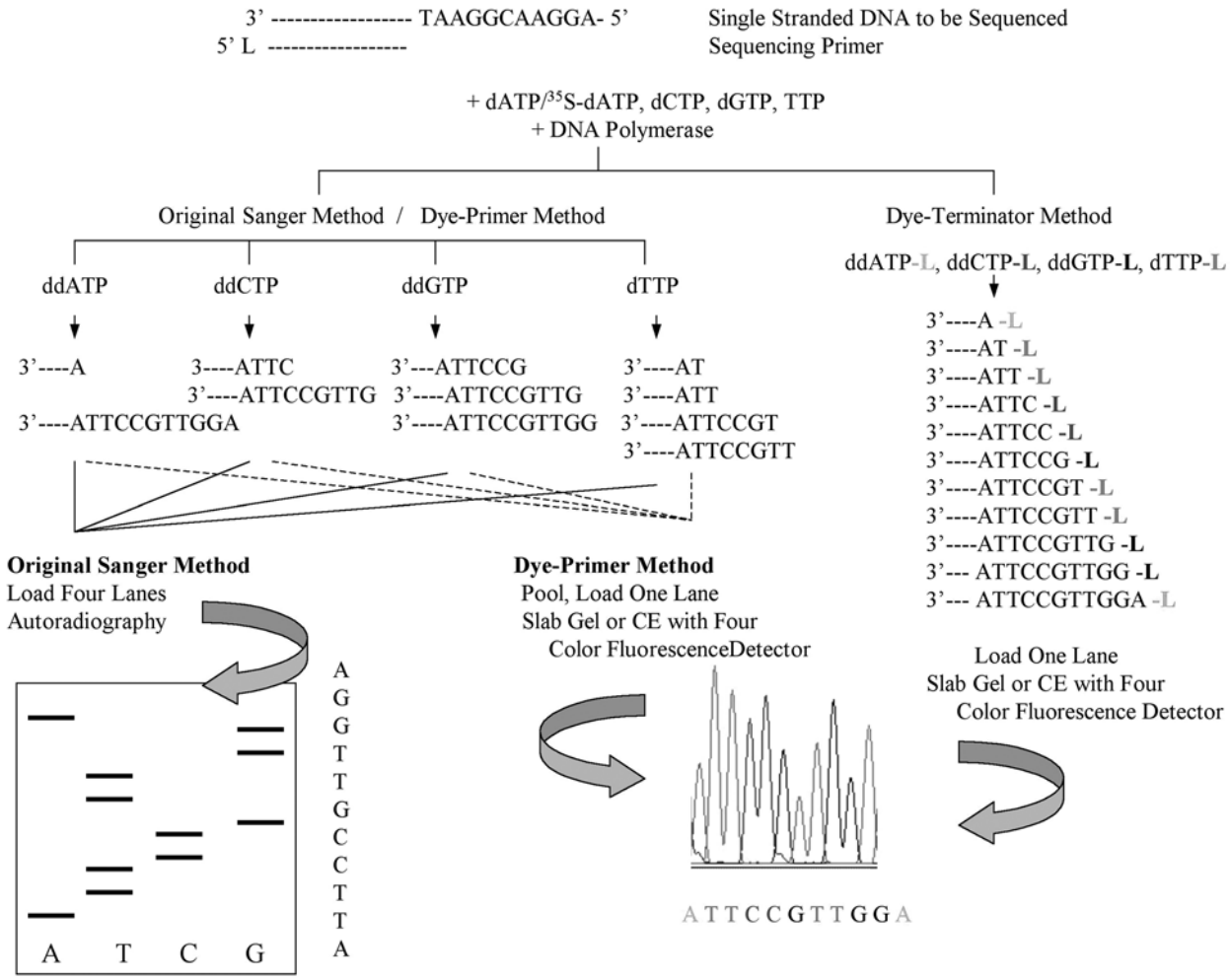


Fig. 18. A schematic of three popular permutations of the Sanger sequencing chemistry. Shown are the original Sanger method in which labeling is accomplished by inclusion of ^{35}S- (or ^{33}P-) dATP. The reaction is done in four parts: the G reactions done in the presence of dideoxy GTP, the A reactions in the presence of ddATP, and so forth. Each reaction is loaded onto one lane of a slab gel. Detection is by autoradiography of the dried gel. Next is a schematic of the fluorescent dye–primer method. In this method, the reaction is again done in four parts: The G reaction is done in the presence of unlabeled ddGTP and a sequencing primer labeled with fluorescent dye 1, the A reaction is done with a primer labeled with dye 2, and so forth. After the extension reaction is completed, the individual reactions are pooled and run on one lane (or one capillary) in a sequence analyzer with a multiwavelength fluorescence detector. The last method is the dye–terminator method, in which the reaction is carried out in a single tube in the presence of all four ddNTPs, with the ddATP labeled with dye 1, the ddGTP labeled with dye 2, and so forth. After the extension reaction, the products are electrophoresed in a single lane (or capillary) and detected with a multiwavelength fluorescence detector.

is in the area of labeling technology. The original Sanger method utilized an ^{35}S-labeled deoxynucleotide triphosphate for labeling of the ddNTP terminated fragments. In this method, a radiolabeled dNTP (typically α-^{35}S dATP or α-^{33}P dATP) is included in the reaction mixture. The labeled dNTP does not terminate the growing chain and is incorporated randomly into the growing DNA chains. After electrophoresis on thin polyacrylamide gels, the gels are dried and exposed to large sheets of X-ray film to visualize the bands. This method is straightforward, does not require expensive equipment, and is still widely practiced today. This method does, however, have some drawbacks in addition to the obvious use of radioactivity. Although the electrophoresis is typically carried out on long (40–50 cm) gels, there is a definite limit to the number of bands that can be resolved on each loading. Under routine conditions, there is very high resolution of lower-molecular-weight fragments that have traveled the longest distance through the gel, and there is less resolution with increasing molecular weight and decreasing distance traveled through the gel. One solution to this problem is to "double load" the gel. In this approach, the four reaction mixtures (each terminated by ddATP, ddCTP, ddGTP, or ddTTP) are loaded in adjacent lanes and subjected to electrophoresis for 1–2 h. The electrophoresis is then paused, the samples are loaded a second time into clean wells, and the electrophoresis resumed. After the electrophoresis is continued for a sufficient time for the lowest-molecular-weight bands in the second load to reach the bottom of the gel, these fragment have migrated off the end in the lanes that were loaded first. In these lanes, the longer fragments are resolved, whereas in the lanes with less electrophoresis time, the shorter fragments are resolved. Using this approach, several hundred bases can be resolved on one gel. Other approaches involve reduction of resolution at the bottom of the gel while retaining resolution on the upper portion. An example of this strategy is the use of

wedge spacers. Spacers, placed between the glass plates before casting the gel, define the shape and thickness of the gel. If the spacers are thicker at the bottom than the top of the gel, the effective voltage (in V/cm) is reduced at the bottom of the gel, leading to decreased mobility and less spacing between shorter fragments. Thus, the gels can be run longer before the smaller fragments migrate off the end, and more bases can be read. The proprietary gel formulation Long Ranger® (Cambrex Bio Science Walkersville, Inc., Walkersville, MD) yields an electrophoretic pattern similar to those seen with wedge spacers, but without the difficulty of pouring non-uniform-shaped gels.

Fluorescent labeling of ddNTP terminated fragments was a major advance in sequencing technology. Smith et al. described a system in 1985 that involved labeling the 5′ end of the sequencing primer with four different fluorescent dyes *(68,69)*. The sequencing reactions were carried out in four aliquots, as for radioactive sequencing. However, in the fluorescent approach, the A reaction was carried out with a primer labeled with one fluor, the C reaction was carried out with a primer labeled with a second fluor, and so on. After the reactions were complete, they were pooled and run together in one lane of a polyacrylamide gel and detected with a multiwavelength fluorescence detector located at a fixed position near the bottom of the gel. Automated sequencing was born. Other approaches utilized primer labeling with one fluorescent dye and keeping the A, C, G, and T reactions separate with electrophoresis in four lanes *(70)*. The multiple-fluor approach increased thoughput by reducing the number of lanes required by a factor of 4, whereas the single-fluor approach decreased instrument complexity. In addition, different geometries for fluorescence excitation and detection have been proposed. Smith et al. utilized a detection format in which the laser and detector are both mounted in a housing that is moved back and forth near the bottom of the gel *(69)*. Ansorge et al. used a number of fixed, single-wavelength detectors and a fixed laser that illuminated the fluors through the gel from the side *(70)*. The four-color, one-lane, moving-detector instrument design was marketed by Applied Biosystems (Foster City, CA) as the 370, 373, and 377 series of DNA sequencers. The one-color, four-lane, fixed-detector design was marketed by Pharmacia (now Amersham Biosciences [Piscataway, NJ]) as the ALF family of sequence analyzers. Both designs have been replaced by CE-based instruments.

The next innovation in sequencing technology was to replace dye–primer chemistry with dye–terminator chemistry. This innovation eliminated the need to prepare an expensive fluor-labeled sequencing primer for each piece of DNA to be sequenced. Perhaps more importantly, it eliminated the need to split each reaction into four. The ability to use unlabeled sequencing primers and the ability to carry out the entire sequencing reaction in one tube allowed for efficient translation to robotic workstations for assay setup and led to a dramatically decreased cost. The four fluorescent dyes that are most commonly used are carboxyfluorescein (abbreviated FAM), carboxy-4′, 5′-dichloro-2′, 7′-dimethoxyfluorescein (JOE), carboxytetramethy-rhodamine (TAMRA), and carboxy-X-rhodamine (ROX). These dyes are particularly useful because their emission wavelengths are evenly spaced, which facilitates discrimination between colors and gives highly accurate base calls. Two lasers are required, however: one at 488 nm for excitation of the fluorescein dyes (FAM and JOE) and one at 543 nm for excitation of the rhodamine dyes (TAMRA and ROX).

The most recent innovation is the introduction of energy transfer dyes (ET dyes) by Mathies and colleagues *(71)*. Fluorescence resonance energy transfer (FRET) is a quantum phenomenon that takes place when two fluorescent molecules (1) have excitation and emission maxima such that the excitation spectra of one fluor overlaps the emission spectra of the other and, (2) the two molecules are very close to one another when the fluor with the lower excitation wavelength is illuminated at its excitation maxima. Under these conditions, when the fluorescent molecule with the lower excitation wavelength moves to the excited state after absorption of a photon, instead of relaxing to the ground state by emission of a photon, a nonradiative transfer of energy occurs in which the first fluor returns to the ground state and the second molecule is excited and fluoresces at its higher emission wavelength. FRET has numerous applications in molecular biology and is the basis for many real-time PCR techniques, such as TaqMan, Molecular Beacons, and FRET probes. As labels for DNA-sequencing reactions, they have the advantage of only requiring a single laser for excitation. The ET dyes use as second fluors the same FAM, JOE, TAMRA, and ROX dyes used for single-dye labeling. They have been adapted to dye–terminator chemistry *(72)* and are marketed by Applied Biosystems (Foster City, CA) under the name BigDye®. The "big dyes" are markedly superior to single dyes in terms of signal strength, lack of differential effects on DNA mobility between dyes, and ability to detect heterozygotes *(73)*. The use of these dyes has virtually replaced the use of the single dyes. The use of ET dye–terminator chemistry combined with cycle sequencing *(74)* and CE separations are the combination of methods of choice for many DNA diagnostics laboratories. An example of sequences traces around the zidovudine (AZT)-resistance mutation at codon 215 of the human immunodeficiency virus (HIV) reverse transcriptase gene using these techniques in a 'home-brew' protocol is shown in Fig.19.

7.3. CAPILLARY ARRAY INSTRUMENTS That CE is superior in many respects to slab gel electrophoresis has been previously noted in this chapter. Other advantages include its speed, a DNA-sequencing run reading 5–600 bases can be completed in approx 2 h, whereas the same separation would typically require 12–16 h on a slab gel instrument. However, slab gels have the clear advantage of being able to handle multiple samples (up to 96 lanes) on 1 gel, whereas CE offers one sample analysis at a time. In 1992, Mathies and Huang introduced the concept of the capillary array electrophoresis (CAE) *(75)*. This system utilized 25 parallel capillaries and injection and detection systems that allowed the parallel analysis of 25 DNA-sequencing reactions. Others soon followed with different methods of accomplishing CAE. Currently, to the author's knowledge, there are 3 manufacturers offering 96 capillary array instruments commercially. The Molecular Dynamics (Sunnyvale, CA) MegaBace 1000 instrument is based on the design of the Mathies group *(75)*. The Applied Biosystems (Foster City, CA) 3700 instrument is based on the work of

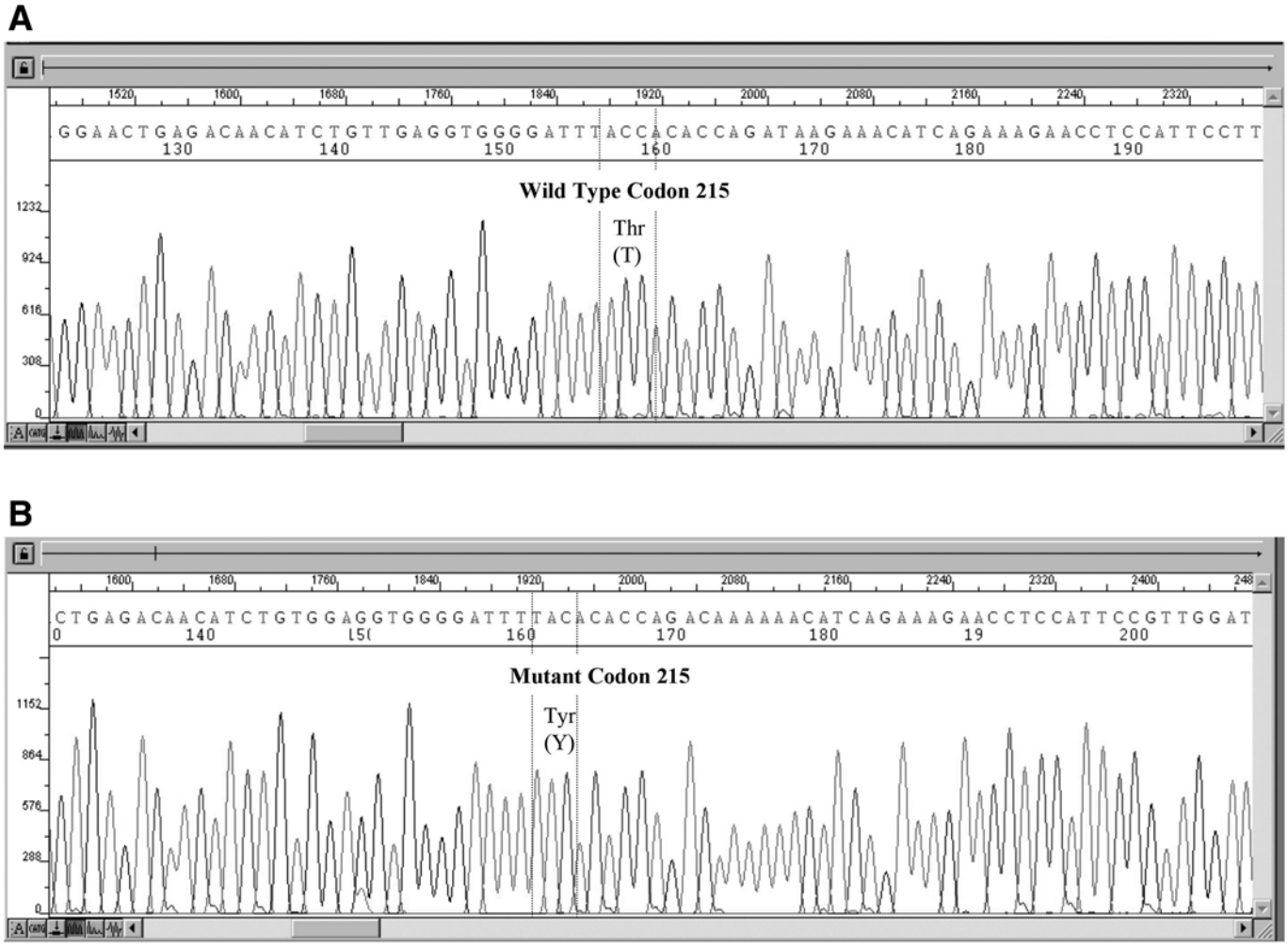


Fig. 19. Partial sequence of the HIV reverse transcriptase gene generated using ET dye–primer chemistry and capillary electrophoresis on an ABI 3100 sequence analyzer (Applied Biosystems, Foster City, CA). Shown are 40-base segments of a sequence from the HIV-1 reverse transcriptase genes from **(A)** a wild-type, AZT-sensitive laboratory strain of HIV-1 and **(B)** a mutant sequence from a patient who failed AZT therapy. Note the presence of the two point mutations resulting in the T215Y amino acid change. (Courtesy of Heui Ra Yoo, University of Maryland Medical Systems.)

Table 2
Comparison of Some Salient Features of Three CAE Instruments

	Molecular Dynamics	*Applied Biosystems*	*SpectruMedix*
Capillaries	Coated with linear polyacrylamide	Bare silica	Bare silica
Matrix	Linear polyacrylamide	Proprietary	Proprietary
Illumination	Confocal microscope Objective; scanning system	Sheath flow postcolumn detection with side illumination	Side illumination; capillaries immersed in refractive-index-matching liquid
Detection	Confocal microscope Objective; four photomultiplier tubes; filter optics	Stationary charge-coupled device camera; grating optics	Stationary charge-coupled device camera.

Source : Modified from refs. *78* and *79.*

Kambara and Tkokashashi *(76)*, and the SpectruMedix (State College, PA) SCE9600 is based on the design of Yeung and Ueno *(77)*. All three manufacturers offer complete sequencing systems, with hardware and software, and each instrument delivers robust, high-quality DNA-sequence data at the rate of at least 500 bases per capillary in under 3 h. A brief comparison of some of the salient features of each instrument is given in Table 2.

8. SOFTWARE FOR DATA ANALYSIS

A discussion of the full range of bioinformatic tools available for sequence analysis, data mining, and genomics is

beyond the scope of this chapter. Comments here will be limited to software for base calling in automated sequencers. In some respects, the discssion is necessarily brief, as the algorithims used by manufacturers of automated sequence analyzers are proprietary. However, Green and colleagues have developed base calling software, *phred*, that is nonproprietary, is well described in the literature, and is the software of choice for many genomics centers and clinical laboratories *(80)*. This software is freely available to academic users.

Computer analysis is required to convert the image of the gel to a base sequence. This analysis has four components. First, the lanes are tracked; that is, a line (straight or curved) is applied to the actual path of the DNA fragments such that peaks can be assigned to a given sample. Using slab gel automated instrumentation, this can be far from a trivial task, because as a result of inhomogeneities in either the gel or electric field, the path of the DNA fragments can be rather far from straight. The introduction of 96-well slab gels made this problem particularly difficult, as the lanes were so narrow and tightly spaced that it was often difficult to determine which trace went with which sample. However, the introduction of CE-based instrumentation has completely eliminated the lane-tracking problem. After the boundaries of the lane have been identified, the thousands of signals from the detector are summed to produce a lane profile, or trace. Trace processing involves deconvolution and smoothing of the signals, noise reduction, and correction for differential dye effects on fragment mobility. The last step is base calling, in which the processed trace is converted into a sequence of bases.

The *phred* software utilizes a four-step procedure for calling bases. The first step is to localize, in each lane, the ideal position of the peaks. Although the positions of the peaks should be able to be known with a high degree of certainty, there are many factors that will cause one or more peaks to exhibit aberrant mobility. One of these factors is the compression artifact, where the sequencing fragments are imperfectly denatured and the 3′ end of the single stranded fragment folds back onto itself, giving a hairpin loop that has higher electrophoretic mobility than perfectly denatured fragments. Fragments with high G+C content tend to give more compressions than A+T-rich fragments. The use of dye–terminator chemistry seems to decrease (but not eliminate) compression artifacts, presumably as a result of steric hindrance to intramolecular base-pairing by the bulky dye moieties *(81)*. Other factors that can cause deviation from ideal peak spacing include polymerase stops, typically caused by secondary structure of the template during the sequencing reaction, and noise introduced by variation in elongation or labeling resulting from degraded or impure reagents.

The second task that the *phred* software accomplishes is to identify all of the peaks in the electropherogram. Third, the existing peaks are mapped onto the idealized pattern. During this step, some peaks will not be assigned. Finally, any unassigned peaks that appear to represent a base (using a rules-based algorithm) are inserted into the sequence.

The accuracy of the *phred* software was compared to the ABI base calling algorithm by comparing base calls from both pieces of software and deriving the error rates for each vs the known reference sequences. *phred* was found to be superior at all places in the electropherogram (short fragments and long) *(80)*.

Clearly, the question of methods of evaluating base calling accuracy is a crucial one. One method, such as the one used by Ewing et al. *(80)* to validate the accuracy of the *phred* software, is to compare the results from any given sequencing run to a reference sequence or a "finished sequence." The disadvantage of this method is that the accuracy measurements might not be available in real time. For example, in *de novo* genome sequencing, it might be some time between the availability of one or two sequencing runs covering a given area and the availability of a finished sequence, defined as a consensus DNA sequence with several-fold depth of coverage. In the clinical laboratory, the areas typically sequenced are well known and the question is whether an observed sequence variation is the result of a potential disease-causing mutation or a sequencing error. Comparison of a newly generated sequences from a patient to a reference sequence is not particularly helpful in making this distinction. Another potential method to estimate errors is to include one or more control sequences in the analysis. Certainly, this approach is virtually mandatory in the clinical laboratory, but it must be kept in mind that error profiles can differ between sample and controls. Green and colleagues have addressed this critical issue by incorporating direct estimates of the probability of base-calling errors into the *phred* software *(82)*. A review of the statistical methods for accomplishing this task is beyond the scope of this chapter and the interested reader is referred to the original literature *(80,82)*. Briefly, four parameters from the trace files were found to be most effective for discriminating errors from correct base calls: peak spacing, the ratio of the amplitude of the largest uncalled peak to the smallest called peak in a window of seven bases, the ratio of the amplitude of the largest uncalled peak to the smallest called peak in a window of three bases, and the peak resolution. Each base is given a quality score, q, which is defined as

$$q = -10 \log_{10}(p)$$

where p is the estimated error probability for that particular basecall. Thus, a base call with a 1 in 100 probability of being incorrect would have a quality score of 20, and a quality score of 40 would indicate an probability of error of 1 in 10,000. Although no consensus has appeared, it would probably be advisable to limit interpretation of clinical sequencing data to regions with quality scores of greater than 30.

The error rate prediction of *phred* was validated by Richterich *(83)* using data from six different large-scale sequencing projects. Not only did he find that the error prediction to agree extraordinarily well with what was found by comparing individual runs to the finished sequence, but, remarkably, the algorithm was insensitive to a number of procedural variations, including dye–terminator vs dye–primer chemistry, DNA preparation methods, types of fluorescent tag, and different sequencing enzymes.

9. APPLICATIONS OF DNA SEQUENCING IN THE CLINICAL LABORATORY

In the area of clinical molecular genetics, the number of targets for sequencing-based mutation analysis is limited only by the number of genes in the human genome and the number of genes associated with human disease. Although some com-

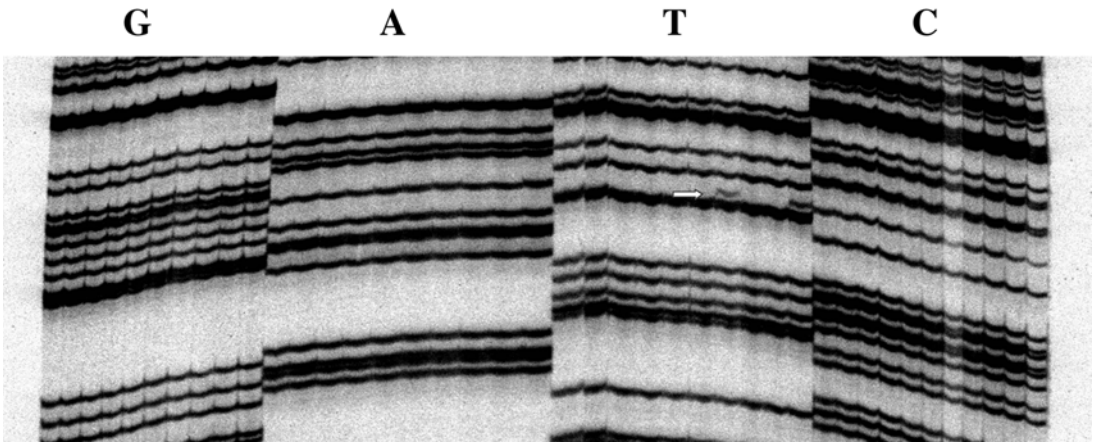


Fig. 20. Slab gel sequencing with radiolabeling run in mutation scanning mode. The autoradiogram shows eight patient samples, a negative control (leftmost lane in each loading) and a positive control, TGC to TAC at codon 609 or C609Y; rightmost lane in each loading). Exon 10 of the RET oncogene was amplified and sequenced using ^{33}P-labeled ddNTPs. All of the G reactions from the 10 samples are loaded in adjacent lanes, then all of the A reactions,and so forth. Note that the presence of mutations (arrow) in the positive control and one of the patients is easily noted by inspection of the autoradiogram.

mercial laboratories and large academic laboratories have offered clinical sequencing for some time for targets with high allelic heterogeneity (e.g., BRCA1 sequencing at Myriad Genetics [Salt Lake City, UT] and CFTR sequencing at Ambry Genetics [Costa Mesa, CA]), other specific or scanning mutation-detection methods are more commonly utilized in the clinical laboratory. However, that situation is rapidly changing as sequencing technology becomes more common in the clinical laboratory. The great majority of DNA-sequencing applications in the clinical laboratory are simply for mutation detection. It is possible to utilize dideoxy sequencing in a "mutation scan" mode, vs a complete sequencing mode. By loading a radiolabeled dideoxy sequencing gel as shown in Fig. 20, it is possible to utilize simple pattern recognition to easily note the presence of a difference between the wild-type control and a patient (or control) bearing a sequence variation. In this case, one harnesses the very high sensitivity and specificity of dideoxy sequencing for the identification, without the detailed analysis required to "read" every A, C, G, and T.

The single test that has introduced DNA sequencing into the majority of molecular diagnostics laboratories had been HIV drug-resistance genotyping, or genotypic antiretroviral resistance testing (GART). The introduction of HAART (highly active antiretroviral therapy), which utilizes a combination of drugs targeting the HIV reverse transcriptase (RT) and protease genes, has transformed HIV infection from a sentence of certain death to a chronic disease; it is one of the most significant therapeutic advances in the past several decades. In the United States and abroad, clinical molecular diagnostics laboratories have been offering HIV viral load assays for several years.

Unfortunately, it is all too common for a patient to start a course of HARRT, have a decrease in the viral load and an increase in the CD4+ cell count, only to relapse, with increasing viral loads, at a later date. GART testing can identify specific changes in the viral genome responsible for acquired resistance to one or more antiretroviral drugs and is essential for tailoring therapy to a patient's specific virus. The evaluation of the genetic composition of HIV has progressed from an epidemiological tool for research into a clinically useful laboratory test that health care professionals might use to select and tailor the appropriate antiretroviral agents to manage HIV disease progression (for a recent review, *see* ref. *84*).

Genotypic antiretroviral resistance testing by DNA sequence analysis and the identification of specific mutations of the RT and protease genes is the most common method of evaluating HIV drug resistance. Those mutations that alter the efficacy of drug binding to its active site are referred to as primary mutations. As a result of primary mutations, higher concentrations of drug are generally required to inhibit the activity of the enzyme. Although viruses with primary drug-resistance mutations can replicate in the presence of the drug, in many cases the fitness of the virus is compromised by these mutations. Secondary mutations are defined as those mutations that supplement the action of the primary mutations by increasing viral fitness. In the absence of primary mutations, secondary mutations generally lack any effect on the level of resistance demonstrated by the virus.

The GART results are usually difficult to interpret. There are several methods of assisting a practitioner with interpreting genotypic information. Tibotech-Virco (Mechelan, Belgium) offers an interpretive service, the Virtual Phenotype™, that is based on a correlative database of more than 100,000 HIV phenotypes and genotypes. Initially, the RT and protease gene sequences of the virus are established with a DNA sequencing-based genotyping assay. The sequence from a given patient is analyzed by computer software that identifies mutations that confer resistance to any of the HAART drugs and then scans the database for similar genotypes from previous samples that might match the patterns of mutations. Once any matches are identified, the phenotypes of these samples are retrieved from the database, and for each drug, an interpretive evaluation is then provided. Although the database was developed using the

Virco genotyping system, the Virtual Phenotype is also available via secure Internet to laboratories who use other genotyping products, including procedures developed in-house by individual laboratories.

In a rules-based approach, the information provided by genotyping is presented as a categorical resistant or sensitive prediction based on a rules-based algorithm. A β-test version of online rules-based interpretation software is currently available at Stanford University's website (http://hivdb.stanford.edu).

There are several analytical platforms currently being marketed for GART testing. Two vendors offer Food and Drug Adminstration-approved systems that use Sanger sequencing on automated DNA sequence analyzers. These systems include PCR amplification reagents, DNA sequencing reagents and hardware, and software that provides a rules-based interpretation of the sequence and generates a user-friendly, easily interpreted report. These systems are available from Visible Genetics (TruGene HIV-1 Genotyping Assay; Visible Genetics, Inc., Toronto, Ontario) and Applied Biosystems (ViroSeq HIV-1 Genotyping System; Applied Biosystems, Foster City, CA). An alternative approach is the line probe assay, available an ASR (analyte specific reagent) from Innogenetics (INNO-LiPA HIV-1 RT, Innogenetics, Inc., Atlanta, GA). This system uses a large set of oligonucleotide probes complementary to known sequence variants associated with drug resistance immobilized on a nylon membrane in a reverse dot-blot format. The advantage of this system is that it has higher sensitivity to minor species, has a much shorter turnaround time, and requires significantly less technical expertise than does DNA sequencing. A disadvantage is that, as new mutations are identified, new strips that incorporate the new probes must be devised, validated, and manufactured.

10. CONCLUSION

Polymerase chain reaction and other molecular techniques have revolutionized analysis of the human genome. This is most striking in the rapid development of techniques that use PCR for the detection of mutations that cause inherited genetic disease. Advances in technology have allowed physicians to confirm difficult clinical diagnoses and to offer prenatal diagnosis and carrier detection, as well as susceptibility testing. In addition, detection of mutations that cause resistance to antiviral drugs is now a cornerstone in the therapy of viral diseases such as HIV and HCV. It is expected that with further developments in automation and technological advances many of the techniques described in this chapter (which are still relatively costly and labor-intensive) will contribute to the routine use of molecular diagnostic testing in clinical practice.

ACKNOWLEDGMENTS

The author gratefully acknowledges the suggestions and comments of Ruth A. Heim, Ph.D., and the generous contribution of figures from Sarah Adkins, Michelle L. Blalock, Kenneth J. Friedman, Ruth A. Heim, Ph.D., and William G. Learning (Molecular Genetics Laboratory, University of North Carolina Hospitals), Lori Terry and J. Carl Barrett, Ph.D. (Laboratory of Molecular Carcinogenesis, National Institute of Environmental Health Science), Marianna M. Goldrick, Ph.D. (Ambion, Inc.), and Heui Ra Yoo (Molecular Diagnostics Laboratory, University of Maryland Medical Systems). The author thanks Kevin Halling, M.D., Ph.D. for his critical review of the manuscript.

REFERENCES

1. Saiki, R. K., Scharf, S., Faloona, F., et al. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230:**1350–1354, 1985.
2. Mullis, K. B. and Faloona, F. A. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods Enzymol.* **155:**335–350, 1987.
3. Brock, T. D. and Freeze, H. *Thermus aquaticus* gen. n. and sp. n., a non-sporulating extreme thermophile. *J. Bacteriol*. **98:**289–297, 1969.
4. Chien, A., Edgar, D. B. and Trela, J. M. Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus. J. Bacteriol.* **127:**1550–1557, 1976.
5. Saiki, R. K., Gelfand, D. H., Stoffel, S., et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239:**487–491, 1988.
6. FMC BioProducts Catalog 1995, Technical Applications, FMC Bio Products, p. 70.
7. Jorgenson, J. W. and Lucas, K. D. Capillary zone electrophoresis. *Science* **222:**266–272, 1983.
8. Tiselius, A. A new apparatus for electrophoretic analysis of colloidal mixtures. *Trans. Faraday Soc.* **33:**524, 1937.
9. Oda, R. P. and Landers, J. P. Introduction to capillary electrophoresis. in *Handbook of Capillary Electrophoresis,* Landers, J. P., ed., CRC, Boca Raton, FL, pp. 1–48, 1997.
10. Ruiz-Martinez, M. C., Berka, J., Belenkii, A., Foret, F., Miller, A. W., and Karger, B. L. DNA sequencing by capillary electrophoresis with replaceable linear polyacrylamide and laser-induced fluorescence detection. *Anal. Chem*. **65:**2851–2858, 1993.
11. Albarghouthi, M. N. and Barron, A. E. Polymeric matrices for DNA sequencing by capillary electrophoresis. *Electrophoresis* **21:** 4096–4111 2000.
12. Albarghouthi, M. N., Buchholz, B. A., Huiberts, P. J., Stein, T. M., and Barron, A. E. Poly-*N*-hydroxyethylacrylamide (polyDuramide): a novel, hydrophilic, self-coating polymer matrix for DNA sequencing by capillary electrophoresis. *Electrophoresis* **23:**1429–1440, 2002.
13. Heller, C. Influence of polymer concentration and polymer composition on capillary electrophoresis of DNA, in *Introduction to the Capillary Electrophoresis of Nucleic Acids: Volume 1,* Mitchelson, K. R. and Cheng, J., Eds. Humana, Totowa, NJ, pp. 111–123, 2000.
14. Chiari, M., Cretich, M., Damin, F., Ceriotto, L., and Consonni, R. New adsorbed coatings for capillary electrophoresis. *Electrophoresis* **21:**909–916, 2000.
15. Orita, M., Suzuki, Y., Sekiya, T., and Hiayashi, K. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5:**874–879, 1989.
16. Hollstein, M., Sidransky, D., Vogelstein, B., and Harris, C. C. *p53* mutations in human cancers. *Science* **253:**49–53, 1991.
17. Levine, A. J., Momand, J., and Finlay, C. A. The *p53* tumor suppressor gene. *Nature* **351:**453–456, 1991.
18. Sheffield, V. C., Beck, J. S., Kwitek, A. E., Sandstrom, D. W., and Stone, E. M. The sensitivity of single-strand conformational polymorphism analysis for the detection of single base substitutions *Genomics* **16:**325–332, 1993.
19. Myers, R. M., Lerman, L. S., and Maniatis, T. Saturation mutagenesis of cloned DNA fragments. *Science* **229:**242–247, 1985.
20. Highsmith, W. E., Jr., Nataraj, A. J., Jin, Q., et al. Use of DNA toolbox for the characterization of mutation scanning methods. II. Evaluation of single-strand conformation polymorphism analysis. *Electrophoresis* **20:**1195–1203, 1999.
21. Lin-Goerke, J. Ye, S., and Highsmith, W. E. Effects of gel matrix on the sensitivity of SSCP analysis: a study of the effects of novel gel matrices, fragment size, GC content, and base alteration. *Am. J. Hum. Genet*. **55** (suppl.):A188, 1994.

22. Ravnik-Glavac, M., Glavac, D., and Dean, M. Sensitivity of SSCP and heteroduplex method for mutation detection in the cystic fibrosis gene. *Hum. Mol. Genet.* **3:**801–807, 1994.
23. Ren, J. High-throughput single-strand conformation polymorphism analysis by capillary electrophoresis. *J. Chromatogr. B* **741:** 115–128, 2000.
24. Kourkine, I. V., Hestekin, C. N., Buchholz, B., and Barron, A. E. High-throughput, high-sensitivity genetic mutation detection by tandem single-strand conformation polymorphism/heteroduplex analysis capillary array electrophoresis. *Anal. Chem*. **74:** 2565–2572, 2002.
25. Atha, D., Wentz, H. M., Morehead, H., Tian, J., and O'Connel, C. D. Detection of p53 point mutations by single strand conformation polymorphism: analysis by capillary electrophoresis. *Electrophoresis* **19:**172–179; 1998.
26. Wentz, H. M., Ramachandra, S., O'Connell, C. D. and Atha, D. Identification of known p53 point mutations by capillary electrophoresis using unique mobility profiles in a blinded study. *Mutat. Res*. **382:**121–132, 1998.
27. Higasa, K., Kukita, Y., Baba, S., and Hayashi, K. Software for machine-independent quantitative interpretation of SSCP in capillary array electrophoresis. *BioTechniques* **33:**1342–1348, 2002.
28. O'Connel, C. D., Atha, D., Oldenburg, M., et al. Detection of p53 gene mutations: analysis by single-strand conformation polymorphism and Cleavase fragment length polymorphism. *Electrophoresis* **20:**1211–1223, 1999.
29. Liechti-Gallati, S., Schnieder, V., Neeser, D., and Kraemer, R. Two buffer PAGE system-based SSCP/HD analysis: a general protocol for rapid and sensitive mutation screening in cystic fibrosis and any other human genetic disease. *Eur. J. Hum. Genet.* **7:** 590–598, 1999.
30. Bhattacharyya, A. and Lilley, D. M. J. The contrasting structures of mismatched DNA sequences containing looped-out bases (bulges) and multiple mismatches (bubbles). *Nucleic. Acids Res*. **17:** 6821–6840, 1989.
31. Keen, J., Lester, D., Inglehaern, C., Curtus, A., and Bhattacharaya, S. Improved detection of heteroduplexes on Hydrolink gels. *Trends Genet*. **7:**5, 1991.
32. Molinari, R. J., Conners, M., and Shorr, R. G. Hydrolink gels for electrophoresis, in *Advances in Electrophoresis*, Volume 6, Chrambach, A., Dunn, M. J., and Radola, B. J., eds., VCH, New York, pp. 44–60, 1993.
33. Highsmith, W. E., Jr., Jin, Q., Nataraj, A. J., et al. Use of a DNA toolbox for the characterization of mutation scanning methods. I. Construction of the toolbox and evaluation of heteroduplex analysis. *Electrophoresis* **20:**1186–1194, 1999.
34. Ganguly, A., Rock, M. J., and Prockop, D. J. Conformation-sensitive gel electrophoresis for rapid detection of single-base differences in double-stranded PCR products and DNA fragments: evidence for solvent-induced bends in DNA heteroduplexes. *Proc. Natl. Acad. Sci. USA* **90:**10325–10329, 1993.
35. Ganguly, A. An update on conformation sensitive gel electrophoresis. *Hum. Mutat.* **19(4):**334–342, 2002.
36. Calladine, C. R., Collins, C. M., Drew, H. R., and Mott, M. R. A study of the electrophoretic mobility of DNA in agarose and polyarylamide gels. *J. Mole. Biol*. **221:**981–1005, 1991.
37. Myers, R. M., Lumelsky, N., Lerman, L. S., and Maniatis, T. Detection of single base substitutions in total genomic DNA. *Nature* **313:**495–498, 1985.
38. Lerman, L. S. and Silverstein, K. Computational simulation of DNA melting and its application to denaturing gradient gel electrophoresis. *Methods Enzymol*. **155:**482–501, 1987.
39. Abrams, E. S., Murdaugh, S. E., and Lerman, L. S. Comprehensive detection of single base changes in human genomic DNA using denaturing gradient gel electrophoresis and a GC clamp *Genomics* **7:**463–475, 1990.
40. Saleeba, J. A., Ramus, S. J., and Cotton, R. G. H. Complete mutation detection using unlabeled chemical cleavage. *Hum. Mutat*. **1:**63–69,1992.
41. Cotton, R. G. H. Cleavage of mismatched based using chemical reagents. in *Mutation Detection: A Practical Approach,* Cotton, R. G. H., Edkins, E., Forrest, S., eds. Oxford University Press, Oxford 1998.
42. Myers, R. M., Larin, Z., and Maniatis, T. Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA : DNA duplexes. *Science* **230:**1242–1246, 1985.
43. Gibbs, R. A., and Caskey, C. T. Identification and localization of mutations at the Lesch–Nyhan locus by ribonuclease A cleavage. *Science* **236:**303–305, 1987.
44. Marini, J. C., Lewis, M. B., Wang, Q., Chen, K. J. and Orrison, B. M. Serine for glycine substitutions in type I collagen in two cases of type IV osteogenesis imperfecta (OI). *J. Biol. Chem.* **268:** 2667–2673, 1993.
45. Forrester, K., Almoguera, C., Han, K., Grizzle, W. E. and Perucho, M. Detection of high incidence of K-*ras* oncogenes during human colon tumorigenesis. *Nature* **327:**298–303, 1987.
46. Murthy, K. K., Shen, S. -H., and Banville D. a sensitive method for detection of mutations: A PCR-based RNase Protection assay. *DNA Cell. Biol*. **14:**87–94, 1995.
47. Goldrick, M. M., Kimball, G. R., Liu, Q., Martin, L. A., Sommer, S. S., and Tseng, J. Y. NIRCA: a rapid robust method for screening for unknown point mutations. *Biotechniques*. **21:**106–112, 1996.
48. Yoshino, K., Nishigaki, K., and Husimi, Y. Temperature sweep gel electrophoresis: a simple method to detect point mutations. *Nucleic Acids Res*. **19:**3153, 1991.
49. Wong, L. J., Liang, M. H., Kwon, H., Park, J., Bai, R. K., and Tan, D.J. Comprehensive scanning of the entire mitochondrial genome for mutations. *Clin. Chem.* **48:**1901–1912, 2002.
50. Cotton, R. G. H., Rodrigues, N. R., and Campbell, R. D. Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxlamine and osmium tetoxide and its application to the study of mutations. *Proc. Natl. Acad. Sci. USA* **85:**4397–4401, 1988.
51. Goldrick, M. M. RNase cleavage-based methods for mutation/SNP detection, past and present. *Hum. Mutat*. **18:**190–204, 2001.
52. Kogan, S. C., Doherty, M., and Gitschier, J. An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. *N. Engl. J. Med.* **317:**985–990, 1987.
53. Feldman, G. L., Williamson, R., Beaudet, A. L., and O'Brien, W. E. Prenatal diagnosis of cystic fibrosis by DNA amplification for detection of KM-19 polymorphism. *Lancet* **8602:**102, 1988.
54. Yuan, R. Structure and mechanism of multifunctional restriction endonucleases. *Annu. Rev. Biochem.* **50:**285–315, 1981.
55. Highsmith, W. E., Burch, L. H., Zhou, Z., et al. Identification of a splice site mutation (2789 + 5G > A) associated with small amounts of normal cystic fibrosis transmembrane conductance regulator mRNA and mild cystic fibrosis. *Hum. Mutat*. **9:**332–338, 1997.
56. Haliassos, A., Chomel, J. C., Tesson, L., et al. Modification of enzymatically amplified DNA for the destruction of point mutations. *Nucleic Acids Res.* **17:**3606, 1989.
57. Friedman, K. J., Highsmith, W. E., and Silverman, L. M. Detecting multiple cystic fibrosis mutations by polymerase chain reaction-mediated site-directed mutagenesis. *Clin. Chem.* **37:**753–755, 1991.
58. Gasparini, P., Bonizzato, A., Dognini, M., and Pignatti, P. F. Restriction site generating–polymerase chain reaction (RG-PCR) for the probeless detection of hidden genetic variation: application to the study of some common cystic fibrosis mutations. *Mol. Cell. Probes* **6:**1–7, 1992.
59. Lindeman, R., Hu, S. P., Volpato, F., and Trent, R. J. Polymerase chain reaction (PCR) mutagenesis enabling rapid non-radioactive detection of common β-thalassaemia mutations in Mediterraneans. *Br. J. Haematol.* **78:**100–104, 1991.
60. Maxam, A. M. and Gilbert W. A new method for sequencing DNA A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* **74:** 560–564, 1977.
61. Sanger, F., Nicklen, S., and Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci USA* **74:** 5463–5467, 1977.
62. Franca, L. T. C., Carriho, E., and Kist, T. B. L. A review of DNA sequencing techniques. *Qu. Rev. Biophys*. **35:**169–200, 2000.

63. Slatko, B. E., Albright, L. M., and Tabor, S. DNA sequencing: commentary. in *Current Protocols in Molecular Biology*, Ausubel, F. M., Brent, R., Kingston, R. E., et al., eds., Wiley, New York, pp. 7.4A. 19–7.4A.39, 1998.
64. Tabor, S. and Richardson, C. C. Selective oxidation of the exonuclease domain of the bacteriophage T7 DNA polymerase. *J. Biol. Chem*. **262:**15,330–15,333, 1987.
65. Tabor, S. and Richardson, C. C. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA* **84:**4767–4771, 1987.
66. Tabor, S. and Richardson, C. C. Effect of manganese ions on the incorporation of dideoxynucleotides by T7 DNA polymerase and *E. coli* DNA polymerase. *Proc. Natl. Acad. Sci. USA* **86:**4076–4080, 1989.
67. Tabor, S. and Richardson, C. C. A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy and dideoxyribonucleotides. *Proc. Natl. Acad. Sci. USA* **92:** 6339–6343, 1995.
68. Smith, L., Fung, S., Hunkapiller, M., Hunkapiller, T., and Hood, L. The synthesis of oligonucleotides containing an aliphatic amino group at the 5′ terminus: synthesis of fluorescent DNA primers for use in DNA sequnce analysis. *Nucleic Acids Res.* **13:**2399–2412, 1985.
69. Smith, L., Sanders, J., Kaiser, R., et al. Fluorescence detection in automated DNA sequence analysis. *Nature* **321:**674–679, 1986.
70. Ansorge, W., Sproat, B., Stegemann, J., Schwager, C., and Zenke, M. Automated DNA sequencing: ultrasensitive detection of fluorescence bands during electrophoresis. *Nucleic Acids Res*. **15:** 4593–4602, 1987.
71. Ju, J., Ruan, C., Fuller, C. W., Glazer, A. N., and Mathies, R. A. Fluorescence energy transfer dye-labeled primers for DNA sequencing and analysis. *Proc. Natl. Acad. Sci. USA* **92:**4347–4351, 1995.
72. Lee, L. G., Spurgeon, S. L., Heiner, C. R., et al. New energy transfer dyes for DNA sequencing. *Nucleic Acids Res*. **25:**2816–2822, 1997.
73. Zakeri, H., Amparo, G., Chen, S. M., Spurgeon, S., and Kwok, P. Y. Peak height pattern in dichlororrhodamine and energy transfer dye terminator sequencing. *BioTechniques* **25:**406–414, 1998.
74. Rosenblum, B. B., Lee, L. G., Spurgeon, S. L., et al. New dye-labeled terminators for improved DNA sequence patterns. *Nucleic Acids Res*. **25:**4500–4504, 1997.
75. Huang, X. C., Quesada, M. A., and Mathies, R. A. DNA sequencing using capillary array electrophoresis. *Anal. Chem*. **64:**2149–2154, 1992.
76. Kambara, H. and Takahashi, S. Multiple-sheathflow capillary array DNA analyser. *Nature* **361:**565–566, 1993.
77. Ueno, K. and Yeung, E. S. Simultaneous monitoring of DNA fragments separated by electrophoresis in a multiplexed array of 100 capillaries. *Anal. Chem*. **66:**1424–1431, 1994.
78. Carrilho, E. DNA sequencing by capillary array electrophoresis and microfabricated array systems. *Electrophoresis* **21:**55–65, 2000.
79. Kheterpal, I. and Mathies, R. A. Capillary array electrophoresis. *Anal. Chem*. **71:**A31–A37, 1999.
80. Ewing, B., Hillier, L., Wendl, M. C., and Green, P. Base-calling of automated sequencer traces using *phred*. I. Accuracy Assesment. *Genome Res*. **8:**175–185, 1998.
81. Lee, L. G., Connell, C. R., Woo, S. L., et al. DNA sequencing with dye-labeled terminators and T7 DNA polymerase: effect of dyes and dNTPs on incorporation of dye-terminators and probability analysis of termination fragments. *Nucleic Acids Res*. **20:**694–699, 1992.
82. Ewing, B. and Green, P. Base-calling of automated sequencer traces using *phred*. II. Error probabilities. *Genome Res*. **8**:186–194, 1998.
83. Richterich, P. Estimation of errors in "raw" DNA sequences: a validation study. *Genome Res*. **8:**251–259, 1998.
84. Wilson, J. W. Update on antiretroviral drug resistance testing: combining laboratory technology with patient care. *AIDS Reader* **13:** 25–30, 2003.

10 Single-Nucleotide Polymorphisms

Testing DNA Variation for Disease Association

ULRICH BROECKEL AND MARTIN J. HESSNER

1. INTRODUCTION

The identification of DNA variation or gene mutations, which contributes to or determines a disease, has been a major research focus, as it describes the underlying inherited disease components. With the human genome being sequenced, the reference sequence is now at hand, facilitating the systematic identification of DNA variation and its subsequent correlation to health and disease. The most common form of genetic variation in the human genome is a simple change of a basepair, a single-nucleotide polymorphism (SNP). In order to elucidate which SNPs might determine a disease as well as diagnostically type those that already have a disease association, there is a substantial need for genotyping assays that are rapid, reliable, and cost-effective. In this chapter, we describe the various areas in which SNPs will be used as well as an overview of some of the most commonly used SNP genotyping methods.

2. IDENTIFICATION OF SINGLE-NUCLEOTIDE POLYMORPHISMS

A variety of methods have been described for the *de novo* identification of SNPs. At present, the most reliable method of identifying SNPs is by direct resequencing of a limited number of individuals. The identification of the exact DNA sequence for each individual allows for the comparison of the DNA sequence and the identification of its subsequent variation *(1)*. In order to reduce the sequencing effort, a number of alternative strategies have been described, such as SSCP (single-sequence conformation polymorphism) or the analysis of pooled DNA samples. SSCP utilizes the fact that the migration speed of a polymerase chain reaction (PCR) product through a gel depends not only on its size but also on its conformation. Therefore, PCR products with a mutation might migrate differently if separated on an appropriate gel. Because SSCP provides indirect evidence for the presence of a SNP, it requires subsequent sequencing of positive PCR products. In comparison to resequencing of a single individual in a single reaction, the pooling of DNA samples offers a method that reduces the cost for SNP identification by reducing the numbers of reactions performed. On the other hand, this strategy has some limitations because one cannot exactly estimate the frequency of every SNP and rare variants might not be detected *(2)*. Furthermore, individual genotypes are not obtained and, therefore, it is not possible to calculate the measurements of the "relatedness" of the genotypes for a number of SNPs, known as linkage disequilibrium.

With the significant interest in SNPs, a number of groups have already started to identify SNPs as part of large-scale projects and the obtained information has been annotated in publicly as well as commercially available databases. At present, over 4 million SNPs have been detected and deposited into the public SNP databases (http://www.ncbi.nlm.nih.gov/SNP). Subsequent analysis of these data has already yielded interesting results with regard to the frequency and pattern of SNPs in the genome. For example, SNPs are observed in the human genome at an estimated frequency of once every 1000 basepairs; however, this can vary depending on the region of the genome *(3)*. The frequency of SNPs varies to some extend between coding and noncoding regions *(4)*. More important, the frequency of SNPs is not uniform over different populations. A comparison of SNPs between Caucasians and African-Americans shows a larger number of SNPs as well as more SNPs with a lower frequency in African-Americans *(5–7)*. This is consistent with the finding that Africans represent an older population in terms of their population history. Because population history, selection, and other population genetics factors determine the frequency and pattern of SNPs, a careful selection of SNPs is required for the selected population when performing disease-association studies.

3. SNPS AND DISEASE

3.1. TESTING FOR DISEASE CAUSING MUTATIONS OR CLINICALLY RELEVANT POLYMORPHISMS As mutations underlying a number of disease have already been identified, SNPs have great importance in many clinical diagnostic settings, including but not exclusive to (1) the extensive diversity of the human leukocyte antigen (HLA) loci, important in solid organ and bone marrow transplantation, (2) the allelic variation found in hematological antigen systems relevant in

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

blood banking and maternal–fetal incompatibilities, as well as (3) the numerous polymorphisms associated with coagulation disorders/cardiovascular disease, such as the Factor V (Leiden) G1691A, Factor II G20210A, or methylene tetrahydrofolate reductase C677T mutations *(8–10)*. In addition, the underlying mutations for a broad variety of rare diseases have been identified. For example, mutations in a number of genes have been identified as the underlying cause for rare forms of hypertension and hypotension *(11–13)*. Although it is still under investigation whether the genes underlying rare disease also contribute to common diseases such as primary hypertension, these described mutations are not routinely used for diagnostic testing. The identification of the mutations underlying cystic fibrosis was a significant scientific finding as an example for more common diseases and a large number of studies followed investigating the effect of mutations in the underlying gene with regard to disease characteristics and progression *(14)*. Most recently, the identification of mutations in CARD15/NOD2 determining the risk of inflammatory bowel disease (IBD) might serve as an additional example of the progress toward the identification of SNPs for more common diseases *(15,16)*. Subsequently, a number of studies examined the effect of these mutations on the disease pattern and clinical characteristics of IBD in different populations *(17–19)*. Although the number of SNPs, which are used as direct diagnostic tests, is still limited, these first examples might demonstrate that the need for diagnostic typing will increase substantially in the future. With regard to the assay requirements for this purpose, diagnostic labs might mainly use a limited number of SNP assays with pre-established methods, which are robust and highly reliable and can be performed without significant specialization in a routine laboratory environment.

3.2. SNPS FOR MAPPING OF DISEASE-CAUSING MUTATIONS The identification of a disease-causing mutation using a mapping and positional cloning approach procedes the use of a SNP as a diagnostic marker. Association studies, the most widely used study design, compare the frequency of a SNP or a number of SNPs in close proximity between a subset of individuals with the disease and a set of healthy control individuals. Assuming that a SNP influences the disease, a difference in the frequency between the two groups will be observed. SNP-based association studies can be conducted in two possible ways: (1) direct testing of a SNP with functional consequence (i.e., SNPs in coding sequence) or (2) using anonymous SNP. The latter approach utilizes the information that SNPs, that are in close proximity to each other are not independent but rather show distinct levels of association between themselves. The nonrandom association of SNPs is termed "linkage disequilibrium" (LD). LD can, therefore, be utilized not only as a test for the association of the typed SNPs but also as a test for the association of all SNPs that are in LD with the actually assayed SNP *(20)*. Subsequently, the extent of LD is a major determinant for how many SNPs might have to be typed. If LD is strong and extends over a larger segment of DNA, only a limited number of markers have to be typed, as the assayed markers are also associated with the neighboring markers. However, if LD is weak, a high density set of SNPs will be necessary to investigate a region of interest. Association studies can be performed using the candidate gene approach whereby SNPs in a particular gene are interrogated based on previous biological knowledge or hypothesized relationship. An alternative approach represents a search in an identified region or even the whole genome whereby no prior hypotheses are assumed. In any case, the extent of LD again determines how many SNPs have to be tested. In the future, high-density SNP maps can make association studies possible, covering large parts of the human genome. A number of projects are now under way in order to determine the extend of LD in genes as well as the genome *(21)*. Although there are arguments that the progress in the hunt for the genes underlying common diseases such as cancer or cardiovascular disease has been slower than anticipated, most recent findings give raise to cautious optimism *(22)*. In any case, the number of different SNPs that have to be typed are large and the requirements for SNP typing assays are, therefore, more complex for this application in comparison to a solely diagnostic typing. Consequently, the key technical issue for the identification of complex diseases is related to the improvement of SNP typing technologies. For example, typing technologies have to be robust and flexible for a broad variety of SNPs performing for various underlying sequence patterns such as GC-rich regions.

4. METHODS FOR SNP TYPING

4.1. TRADITIONAL APPROACHES TO SNP TYPING Historically, a number of different SNP typing approaches have been used in both research and clinical laboratories. One of the first approaches for SNP typing was polymerase chain reaction–restriction fragment length polymorphism (PCR-RFLP), which can be used for genotyping when a SNP creates or abolishes a restriction site. The technique is performed by amplifying the region of interest with flanking primers, followed by digestion of the PCR product with the appropriate enzyme and resolving of the products by gel electrophoresis. This approach has been commonly used *(23–26)*, even in instances where the SNP does not alter a restriction pattern through the use of PCR-mediated site-directed mutagenesis *(27–31)*. This derivative of traditional PCR-RFLP uses primer modification along with the SNP to create/abolish a restriction site. The primary weaknesses of this approach are its labor–intensiveness and the fact that no or incomplete digestion of the PCR product can lead to erroneous results; however, additional endogenous or engineered restriction sites within the product can serve as internal controls for digestion. Allele-specific PCR (ASPCR), also known by amplification refractory mutation system (ARMS) or PCR with sequence-specific primers (PCR-SSP), is based on the fact that *Taq* polymerase will less efficiently extend an oligonucleotide primer when the 3′ end of a priming oligonucleotide is not perfectly matched with its target template *(32–34)*. As a result, primers can be designed that discriminate single-base substitutions, small deletions, or insertions. Assays are typically arranged in a two-reaction format for biallelic systems where each reaction possesses primers specific for detection of a single allele. The assay end point is typically electrophoretic analysis of the reaction products, and detection of the appropriately sized product is indicative of allelic presence.

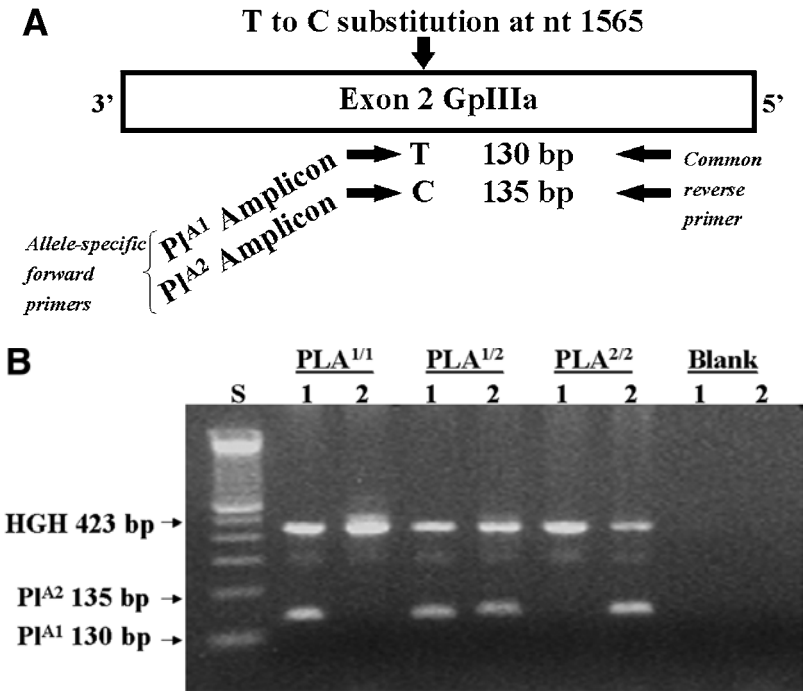


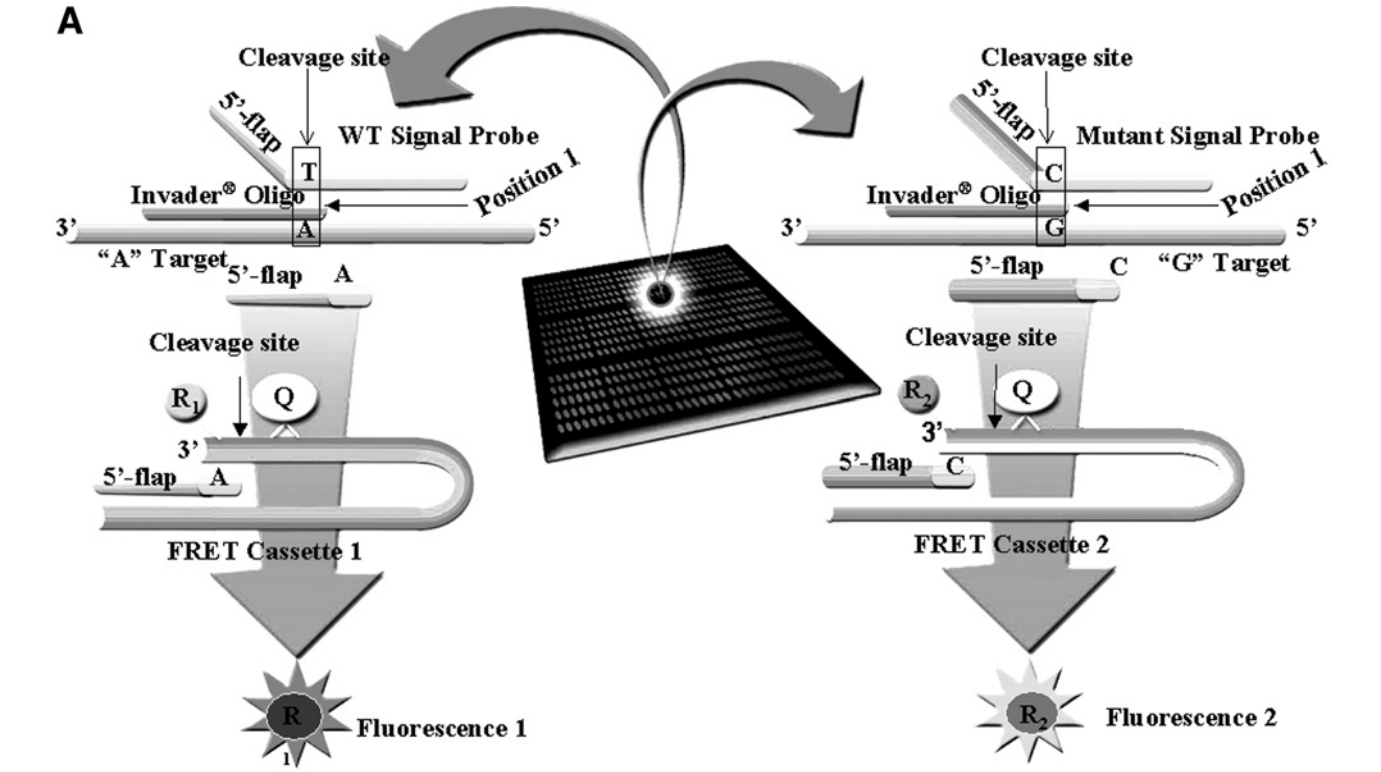
Fig. 1. Genotyping the T1565C ($Pl^{A1/A2}$) dimorphism of the GPIIIa (platelet fibrinogen receptor) gene by allele-specific PCR. **(A)** The two-reaction format: Allele-specific forward primers are paired with a common reverse primer for amplification of either the Pl^{A1} allele (T at 1565) or Pl^{A2} allele (C at position 1565). A second pair of primers, specific for the human growth hormone locus, are included in each reaction as an internal positive control. After amplification products of each analysis are run in pairs **(B)**, the Pl^{A1} reaction is in the first lane of each pair and the Pl^{A2} reaction is in the second lane of each pair and Product sizes and genotypes of illustrated examples are indicated. Lane S: 100-bp DNA ladder (Gibco-BRL Gaithersburg, MD).

Coamplification of a second (uniquely sized) product is often used as an internal control, so that technical failure is not confused with absence of the allele (Fig. 1A,B). The specificity of the allele-specific primers can be enhanced through the introduction of additional primer–template mismatches at the -3, -2, or penultimate nucleotide; this additional destabilization of the 3′ end of the primer reduces mispriming of the genomic DNA target during the early cycles of the PCR reaction *(33,35)*. Application for this typing approach was first reported for detection of the A to T transversion of the human β-globin gene responsible for sickle cell anemia *(34)*. Since then, this approach has been broadly applied in both the research and clinical laboratory for typing of hematologic antigen systems *(36–43)*, mutations associated with thrombotic or cardiovascular risk *(35,38,44–48)*, HLA typing *(49–54)*, α_1-anitrypsin *(32)*, and steroid 21-hydroxylase *(55)*. The advantage of this approach is that it is less labor-intensive than PCR-RFLP in that no postamplification enzymatic digestion is required; however, considerable reaction optimization can be required. More sophisticated derivatives of this approach have been reported, including multiplexed allele-specific assays *(38,56)* and approaches whereby the allele specific primer(s) are "tailed" with priming sequences for subsequent resolution of polymorphisms located within the initial PCR product, a strategy useful for genotyping highly complex loci possessing multiple polymorphic sites, such as HLA *(57)*.

Presently, there is no single SNP genotyping method that is ideal for all research and/or clinical laboratory applications. Commonly used current chemistries include hybridization, primer extension, and cleavage methods coupled to various detection systems. Numerous biotechnology companies have developed elegant and innovative high-throughput systems combining various allelic discrimination biochemistries, reaction formats, and detection methods. In the following subsections are described a number of commercial systems that have found acceptance in the clinical laboratory setting.

4.2. INVADER™ ASSAY The Invader™ assay is a mutation-detection technology developed by Third Wave Technologies, Inc. *(58–63)*. This microtiter plate-formatted assay uses fluorescence resonance energy transfer (FRET) detection and does not require PCR, restriction digestion, or gel electrophoresis. The assay is based on a novel linear signal amplification technology whereby biallelic SNP systems can be genotyped in a single biplex reaction format.

The technology relies on the specificity of Cleavase® enzymes, a class of naturally occurring and engineered enzymes that recognize and cleave structures that form when the 3′ end of an upstream "invading" oligonucleotide overlaps the hybridization site of the 5′ end of a downstream oligonucleotide probe by at least 1 basepair *(58,59,64,65)*. This activity enables detection of single-nucleotide mismatches immediately upstream of the cleavage site on the downstream DNA strand because mispairing results in the formation of a noninvasive structure that the enzyme does not recognize as a cleavable substrate. The generation of the proper enzyme substrate is dependent on base-pairing at a critical position between the two oligonucleotides and the target nucleic acid, which provides the ability to discriminate single-base changes.



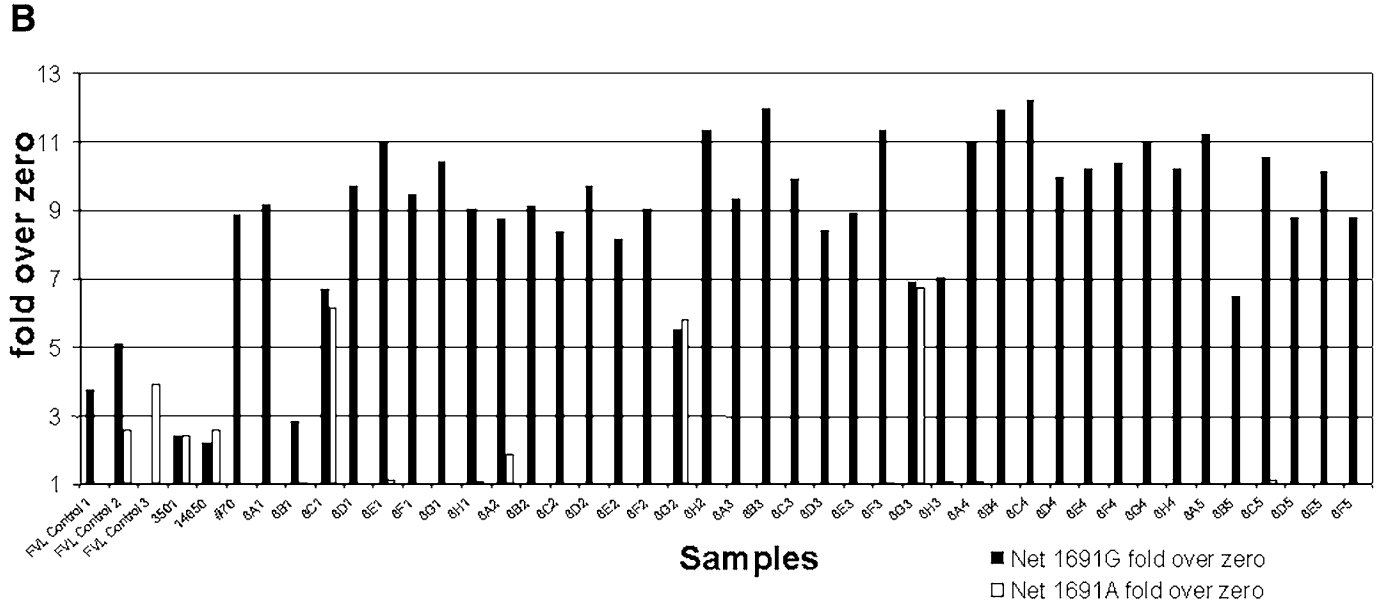


Fig. 2. The Invader assay. **(A)** Assay schematic: This technology system utilizes a 5′ "invading" oligonucleotide and a partially overlapping 3′ "signal" oligonucleotide, which together form a specific structure when bound to a complementary genomic DNA template. A thermostable flap endonuclease cleaves this structure, releasing the 5′ flap from the "signal" oligonucleotide. Elevated temperature and an excess of the signal probe enable multiple probes to be cleaved for each target sequence present without temperature cycling. The cleaved probes then direct cleavage of a secondary probe (FRET cassette) which is 5′-end labeled with a fluorescent dye but is quenched by an internal dye. Upon cleavage, fluorescence is detected using a standard fluorescence plate reader. Using separate 5′ flaps and FRET cassettes possessing different fluorescent dye pairs, detection of two alleles in a single well are possible. Genotypes are determined by through calculation of signal ratio between the two alleles. **(B)** Factor V G1691A (Leiden) Invader genotyping results. For each sample, the "fold-over zero" (FOZ) signal for each probe was determined by dividing the obtained signal from the appropriate "no target blank." Determination of the sample genotypes was based on the ratio of the normal FOZ to the mutant FOZ as follows: Ratio = Normal FOZ/Mutant FOZ. The following ratio values are used to interpret assay results: ≥ 5.0 = normal; ≥ 0.5 to <2 = heterozygous; <0.2 = homozygous mutant; >0.2 to <0.5 *and* >2.0 to 5.0 are equivocal and require resolution through repeating sample. (Figure 2**A** appears in clear in accompanying CD Rom.)

Figure 2A illustrates how the 3′ base of the upstream "Invader oligonucleotide" overlaps with the target-specific region of the "primary probe" (wild-type [WT] or mutant probe) at the base referred to as "position 1." The specificity of the Cleavase enzyme requires that position 1 of the primary probe be complementary to the target for cleavage to occur. A noncomplementary base in the target at position 1 does not lead to formation of an invasive structure. Instead, a nicked structure is created, and position 1 of the primary probe becomes part of a flap that does not hybridize to the target. The Cleavase enzyme does not recognize and cleave the nicked structure. The sequence of the 5′ flap varies in length (typically between 1 and 15 nucleotides), and because it is independent of the target, it can be any sequence. Discrimination relies on enzymatic recognition of the properly assembled structure in addition to the hybridization thermodynamics of oligonucleotides complementary to the target sequence.

When the complementary sequence is present, the Invader oligonucleotide and target-specific region of the primary probe hybridize to the target and form an invasive structure. The Cleavase enzyme removes the 5′ flap from the primary probe. Once released, the 5′ flap forms another invasive structure with a synthetic molecule containing a donor and a quencher fluorophore on each side of the cleavage site (termed "FRET oligonucleotide"). The Cleavase enzyme separates the donor fluorophore from the quencher fluorophore, generating a fluorescent signal. Using two different 5′-flap sequences and their matching FRET oligonueotides with nonoverlapping fluorophores allows for two alleles to be detected in a single well (Fig. 2A). Both reactions (cleavage of the primary probes and of the FRET probes) occur simultaneously at a single temperature near the melting temperature of the primary probe. Thus, numerous primary probes and FRET oligos could be cleaved for each target molecule present, resulting in a linear amplification of signal, but not target. Each step of the assay provides approx 5000 cleavage products per target per hour—over 1-million-fold signal amplification without temperature cycling *(63)*.

The Invader assay can be carried out in microtiter plates, requires minimal hands-on time, and does not require any separation steps. Invader reactions are run for 3–5 h at 63° C, and the end-point fluorescence signal is detected using a standard fluorescence plate reader. Genotypes for homozygous and heterozygous samples are assigned by determining a net wild-type/mutant signal ratio for each sample (Fig. 2B) using Microsoft Excel-based worksheets. This assay platform has been used for the detection of thousands of different SNPs in research settings *(66–71)* and a number of "Analyte-Specific Reagent Kits" are available for typing a number of clinically relevant SNPs, including Factor V (Leiden G1691A), prothrombin G20210A, MTHFR C677T, GP1A T1565C (PI^{A1}/PI^{A2}), and others. These kits have shown high concordance with genotyping generated through more traditional methods *(47,72–74)*.

4.3. READIT® ASSAY Genotyping System (Promega Corporation, Madison, WI) (Fig.3) The READIT® SNP is a three-step system that interrogates PCR amplified samples for a target sequence and produces a bioluminescent light signal to identify which target sequence is present *(75)*. The assay uses hybridization specificity and a coupled reaction using two thermostable enzymes (READase™ Polymerase and READase™ Kinase) to generate high-energy adenosine triphosphate (ATP). When luciferase/luciferin reagent is added, the ATP is used to produce light detectable with a luminometer.

The READIT technology, capable of detecting SNPs, insertions, deletions, and chromosomal translocations, utilizes the ability of DNA polymerases to catalyze hybridization-dependent pyrophosphorolysis *(75–78)*. Pyrophosphorolysis is the addition of inorganic pyrophosphate across the terminal phosphodiester bond, resulting in the shortening of the 3′ end of the DNA substrate and the release of dNTPs. In the case of interrogation of a sample having a biallelic genome, one READIT reaction contains an oligonucleotide probe that can hybridize to one of the known alleles that might be present at the variable site. A second separate but concurrent READIT reaction contains a second oligonucleotide probe that can hybridize to the other known allele. A third separate but concurrent reaction is performed without any added probe to assess background signal from the template. As such, three wells of plate are used to determine each genotype. If the probe hybridizes to the target DNA without a mismatched base within the final six bases at the 3′ end of the probe, an enzyme designated the READase polymerase will carry out a pyrophosphorylation reaction on the primer, generating high-energy dNTPs. If there is a mismatched base within the last bases at the 3′ end of the probe, the READase polymerase will not generate dNTPs from the primer. If dNTPs are generated, a second enzyme present in the reaction (the READase Kinase) will transfer the terminal phosphate from the dNTP to ADP present in the reaction buffer to form ATP. ATP is an essential cofactor in a reaction with luciferase and luciferin that produces light, and the light is used to monitor ATP production. The amount of ATP produced by the reactions is then measured in a second incubation as a light signal generated by a luciferase/luciferin reaction. If a sample is homozygous for one of the alleles, dNTP, then ATP will only be generated in the reaction using the probe that matches that allele. However, if a sample is heterozygous for the alleles in question, essentially equal light signals will be seen in the reactions with both probes.

In order to "clean up" a PCR product prior to READIT assay molecular interrogation, the double-stranded DNA is subjected to enzymatic processing to remove one of two strands of the PCR product as well as to remove residual dNTPs from the spent PCR reactions that would contribute to background in the READIT assay. To protect the strand of the PCR product to be interrogated, one of the two PCR primers must contain three phosphorothioate linkages at the 5′ end of the primer. This modified linkage prevents digestion of this strand during exonuclease treatment. Although either strand of a double-stranded PCR product can be interrogated, it is critical that the READIT Interrogation Probes be complementary to the strand containing the modified primer. Software to assist in READIT Interrogation Probe design is available on the Internet (www.promega.com/readit).

A computer program, the READIT calculator, is available to automatically determine the genotypes of samples analyzed

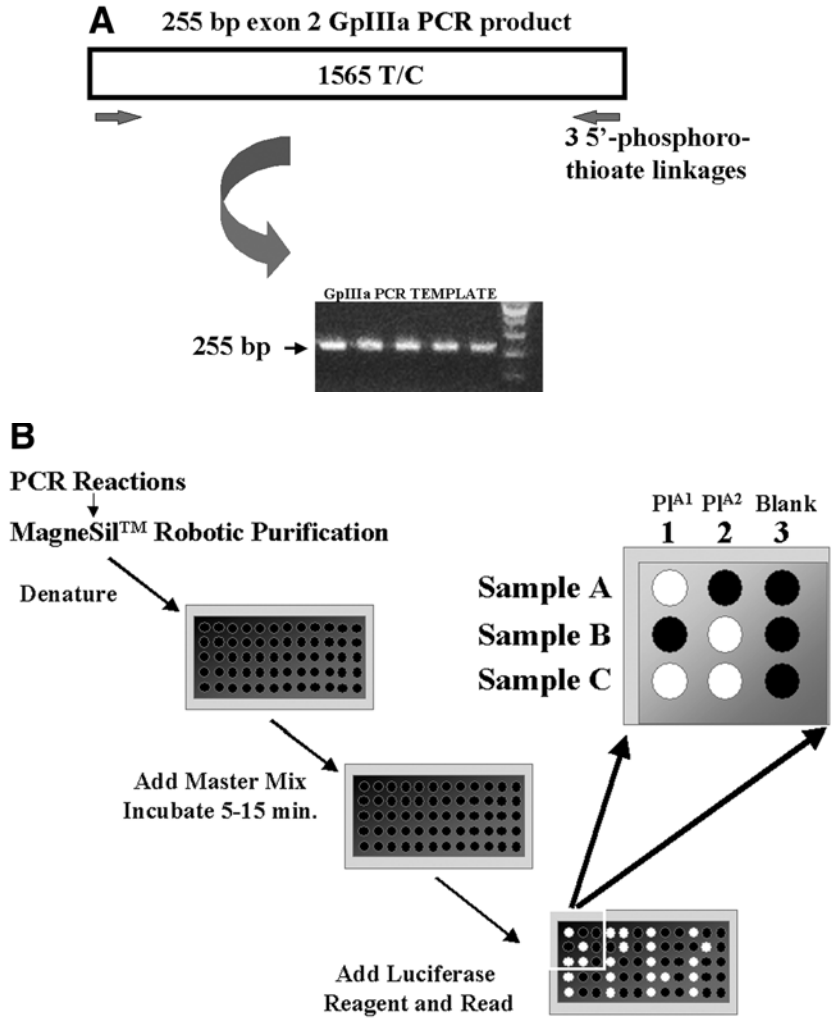


Fig. 3. The READIT assay from Promega Corporation (Madison, WI) utilizes the fact DNA polymerases catalyze hybridization-dependent pyrophosphorolysis, which is the addition of inorganic pyrophosphate across the terminal phosphodiester bond. This results in the shortening of the 3′ end of a DNA probe and the release of dNTPs, when there is recognition between the probe and the target. The analysis of a biallelic polymorphism, such as GPIIIa Pl^{A1}/Pl^{A2}, begins with amplification of the locus. **(A)** where one primer of the primer pair is 5′-modified with 3 phosphorothioate linkages. A single-stranded DNA template is generated by treating the PCR product with exonuclease; the 5′ phosphorothioate linkages protect the READIT™ substrate strand from degradation. A single-stranded DNA template is purified from PCR components and digestion products using Promega's MagneSil™ System. Single-stranded DNA template is interrogated in three reactions. **(B)** In the first reaction, a Pl^{A1}-specific probe is positioned such that the 3′end corresponds to 1565T; likewise, the second reaction possesses a Pl^{A2}-specific probe positioned such that the 3′ end corresponds to 1565C. **(C)** The READIT reaction is conducted at an elevated temperature and only perfect hybrids will serve as substrates for DNA polymerase-mediated pyrophosphorolysis resulting in depolymerization of the interrogation oligonucleotide and the liberation of dNTPs. A third reaction lacks interrogation probe and serves as a background control. A kinase present in the READIT reaction is responsible for randomizing the γ- phosphate present on the released dNTPs onto ADP present in the reaction to produce ATP. **(D)** Once generated, ATP is utilized by luciferase to generate light in reactions where hybridization has occurred. The READIT assay system can be conducted in a 96-well format on robotic workstations for high-throughput applications.

using the READIT system. This program also can be used to track the performance of the system over time and can calculate the population frequencies of the various genotypes analyzed.

4.4. MICROCHIPS (AFFYMETRIX) Affymetrix has developed an interesting technology that utilizes the hybridization of DNA onto a solid support bound oligonucleotide, enabling the simultaneous genotyping of a large number of SNPs. Although the chip technology can be utilized also for expression analysis, it can also be used for SNP genotyping. The methodology is based on an array of oligonucleotides, that are synthesized on glass using a process utilizing photolithographic and combinatorial chemistry. Similar to a computer chip, a wafer serves as the basis, which is subsequently placed in a silane bath. The addition of a linker allows in the subsequent step the activation of the surface by ultraviolet (UV) light. Positions on the chip, which are exposed to UV light become available for the addition of nucleotides in the subsequent steps. Different oligonucleotides can now be synthesized by applying a mask with different windows over the waver, resulting in a selective activation of distinct positions on the waver. After the desired positions on the chip are activated, a solution containing the desired single nucleotide is used to extend the oligonucleotide. A critical step in this manufacturing process is the accurate activation of positions on

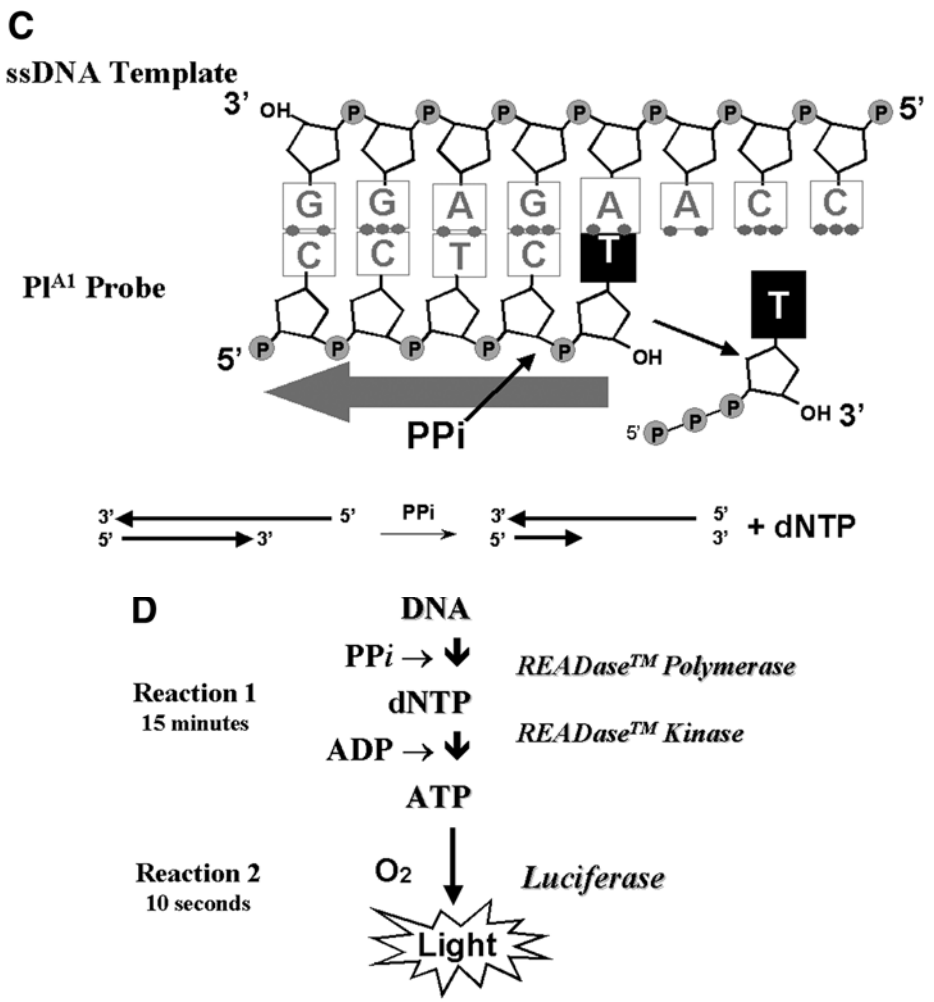


Fig. 3. *(Continued).*

the chip and the accurate attachment of the nucleotides. For SNP genotyping, oligonucleotides are synthesized that are complementary to the SNP of interest with all four different nucleotides at the respective SNP position. SNP typing will then be performed by hybridization of PCR products onto the chip. For the hybridization as well as the measurement and scoring, a workstation provided by Affymetrix is necessary to perform the analysis. The most expensive step for the design of these chips is the manufacturing of the chip-specific masks. Therefore, the strength of method might not be fully utilized for the analysis of a small number of custom SNPs. The strength of this technology might rather be in the simultaneous analysis of a panel or a large number of SNPs. Although custom chips are available (e.g., the CYP polymorphisms), this technology could be utilized with a fixed set of SNP for whole-genome SNP association analysis in a research setting. In addition, for diagnostic purposes, custom chips could be designed that which analyze a larger set of diagnostic SNPs as part of a comprehensive SNP screen.

4.5. TAQMAN The TaqMan assay *(Applied Biosystems, Foster City, CA) (Fig.4)* utilizes the same biochemical approach as been implemented in quantitative PCR analysis. The methodology that is available from Applied Biosystems uses energy transfer as it detects fluorescence based on the physical distance between a fluorophore and a quencher molecule. Allele-specific probes are designed for both alleles of a SNP with the nucleotide of interested located in the middle. The oligo is labeled on the 5′ side with a fluorophore, whereas on the 3′ prime end, a quencher molecule is added. The fluorescence of the 5′ dye is quenched as long as the fluorophore is in close proximity to the quencher. During the PCR reaction, which is performed using flanking primers, the allele-specific probe hybridizes during the annealing step to the PCR product. In the subsequent extension step, using the 5′–3′ exonuclease activity of the polymerase will degrade a matching and annealed probe. This dissociated the fluorophore and the quencher and, subsequently, the fluorescence can be measured. With two allele-specific probes possessing different fluorophores, this reaction allows for the allele discrimination of the SNP. This method has a number of advantages that make it attractive for SNP typing for both diagnostic as well as research applications. The reaction itself is performed in one reaction (PCR and hybridization), requires no additional sample processing, and can be performed in one step without any post-PCR processing. Although assays have to be designed for each SNP individually, this method is most efficient when a limited number of SNPS have to be typed in a large set of samples. This is also because the allele-specific probes are modified with the quencher and fluorophore. Therefore, the increase in the

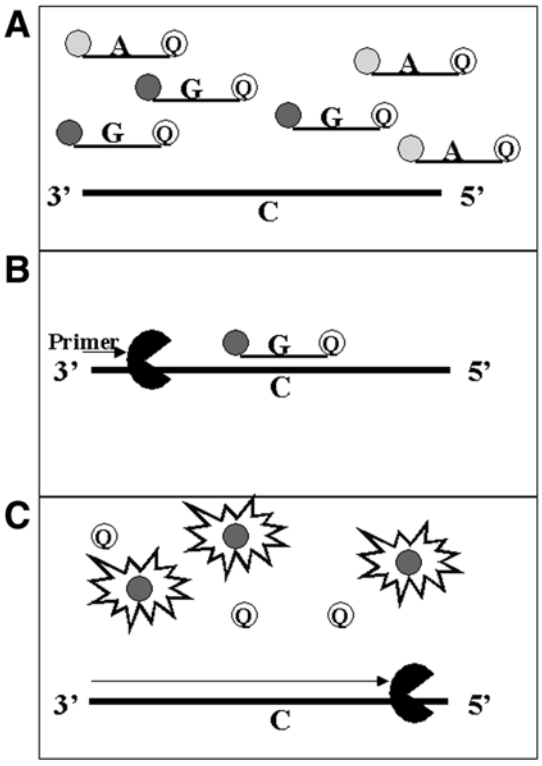


Fig. 4. The TaqMan assay (Applied Biosystems) utilizes the 5′-exonuclease activity of DNA polymerases. The TaqMan assay is homogeneous in that it possesses in each reaction allele-specific probes that are labeled with different fluorophores at their 3′ ends and a quencher molecule at their 5′ ends. **(A)**. The quencher, through FRET, quenches fluorescence as long as the probe molecule is intact. During the annealing phase of PCR, the probes hybridize to their appropriate PCR product target strand, and during the extension phase of PCR, the 5′ exonuclease activity of the *Taq* polymerase degrades perfectly matched and annealed probes **(B)**. The digested probe is released into solution and the fluorophore and quencher are separated, leading to an allele-specific increase in fluorescence with each cycle of PCR **(C)**. Mismatched probes are displaced from the PCR product target strand without digestion.

cost of the probes represents also a significant factor making this method particularly suitable for projects with higher numbers of genotypes. Most recently, predesigned assays are available (assay on demand ™). This set of over 200,000 SNPs with predetermined allele frequency in selected populations can be used for mapping using association analysis in a research setting as well as for diagnostic analysis if the mutation of interest is available.

5. OUTLOOK: NOVEL METHODS AND REQUIREMENTS FOR SEQUENCE-BASED MEDICINE

The methods described in this chapter provide only a limited overview for the numerous SNP typing methods available because new technologies are rapidly evolving. Taken together, the strengths as well as limitations for each assay relate to the ease of establishing an assay, the ability to either type a large number of samples with a limited set of SNPs or a large number of diverse SNPs, and, most importantly, the cost of genotyping. With the technology available as of today, there might not be one single platform, that will fit the need for the broad range of applications in both a diagnostic as well as a research setting. Although sequencing is still the standard for the identification of novel SNPs, new technologies are emerging that potentially reduce the cost and effort of sequencing through a reduction of reaction volumes and automation. A reduction in cost and time might also be achieved in assays, which do not require PCR. Ultimately, one can envision that for research applications, potentially hundreds of thousands of SNPs might be typed in order to perform genomewide association studies. For these applications, highest-throughput methods will be needed and further improvements reducing the cost will be necessary. As more and more disease-causing mutations are identified, one might perform in a diagnostic setting an analysis with a set of a few hundred or thousand SNPs with demonstrated clinical relevance. Ultimately, the ability to assay a whole genome for its polymorphic structure by direct resequencing of each individuals genome might be the most comprehensive analysis. So far, for only a few diseases, the genetic information is being used for diagnosis or a decision on the treatment options. With the focus on the improvement of SNP typing and sequencing, well-designed clinical and epidemiological studies are ultimately the key to correlate genetic variation with disease, treatment, and prevention.

REFERENCES

1. Nickerson, D. A., Tobe, V. O., and Taylor, S. L. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res* **25:**2745–2751, 1997.
2. Sasaki, T., Tahira, T., Suzuki, A., et al. Precise estimation of allele frequencies of single-nucleotide polymorphisms by a quantitative SSCP analysis of pooled DNA. *Am. J. Hum. Genet.* **68:**214–218, 2001.
3. Kwok, P. Y., Deng, Q., Zakeri, H., Taylor, S. L., and Nickerson, D. A. Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. *Genomics* **31:**123–126, 1996.
4. Halushka, M. K., Fan, J. B., Bentley, K., et al. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat. Genet.* **22:**239–47, 1999.
5. Lazarus, R., Klimecki, W. T., Palmer, L. J., et al. Single-nucleotide polymorphisms in the interleukin-10 gene: differences in frequencies, linkage disequilibrium patterns, and haplotypes in three United States ethnic groups. *Genomics* **80:**223–228, 2002.
6. Nickerson, D. A., Taylor, S. L., Weiss, K. M., et al. DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nat. Genet.* **19:**233–240, 1998.
7. Rieder, M. J., Taylor, S. L., Clark, A. G., and Nickerson, D. A. Sequence variation in the human angiotensin converting enzyme. *Nat. Genet.* **22:**59–62, 1999.
8. Bertina, R. M., Koeleman, B. P., Koster, T., et al. Mutation in blood coagulation factor V associated with resistance to activated protein C. *Nature* **369:**64–67, 1994.
9. Bertina, R. M., Reitsma, P. H., Rosendaal, F. R., and Vandenbroucke, J. P. Resistance to activated protein C and factor V Leiden as risk factors for venous thrombosis. *Thromb. Haemost.* **74:**449–453, 1995.
10. Poort, S. R., Rosendaal, F. R., Reitsma, P. H., and Bertina, R. M. A common genetic variation in the 3′-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. *Blood* **88:**3698–3703, 1996.
11. Geller, D. S., Rodriguez-Soriano, J., Vallo Boado, A., et al. Mutations in the mineralocorticoid receptor gene cause autosomal dominant pseudohypoaldosteronism type I. *Nat. Genet.* **19:**279–281, 1998.

12. Lifton, R. P. Molecular genetics of human blood pressure variation. *Science* **272:**676–680, 1996.
13. Simon, D. B., Bindra, R. S., Mansfield, T. A., et al. Mutations in the chloride channel gene, CLCNKB, cause Bartter's syndrome type III. *Nat. Genet.* **17:**171–178, 1997.
14. Collins, F. S. Cystic fibrosis: molecular biology and therapeutic implications. *Science* **256:**774–779, 1992.
15. Hugot, J. P., Chamaillard, M., Zouali, H., et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411:**599–603, 2001.
16. Ogura, Y., Bonen, D. K., Inohara, N., et al. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411:**603–606, 2001.
17. Abreu, M. T., Taylor, K. D., Lin, Y. C., et al. Mutations in NOD2 are associated with fibrostenosing disease in patients with Crohn's disease. *Gastroenterology* **123:**679–688, 2002.
18. Helio, T., Halme, L., Lappalainen, M., et al. CARD15/NOD2 gene variants are associated with familially occurring and complicated forms of Crohn's disease. *Gut* **52:**558–562, 2003.
19. Louis, E., Michel, V., Hugot, J. P., et al. Early development of stricturing or penetrating pattern in Crohn's disease is influenced by disease location, number of flares, and smoking but not by NOD2/CARD15 genotype. *Gut* **52:**552–557, 2003.
20. Cardon, L. R. and Bell, J. I. Association study designs for complex diseases. *Nat. Rev. Genet.* **2:**91–99, 2001.
21. Couzin, J. Human genome. HapMap launched with pledges of $100 million. *Science* **298:**941–942, 2002.
22. Korstanje, R. and Paigen, B. From QTL to gene: the harvest begins. *Nat. Genet.* **31:**235–236, 2002.
23. Cutting, G. R., Kasch, L. M., Rosenstein, B. J., et al. A cluster of cystic fibrosis mutations in the first nucleotide-binding fold of the cystic fibrosis conductance regulator protein. *Nature* **346:**366–369, 1990.
24. Highsmith, W. E., Jr., Burch, L. H., Zhou, Z., et al. Identification of a splice site mutation (2789 + 5 G > A) associated with small amounts of normal CFTR mRNA and mild cystic fibrosis. *Hum. Mutat.* **9:**332–338, 1997.
25. Rootwelt, H., Berger, R., Gray, G., et al. Novel splice, missense, and nonsense mutations in the fumarylacetoacetase gene causing tyrosinemia type 1. *Am. J. Hum. Genet*. **55:**653–658, 1994.
26. Vohl, M. C., Couture, P., Moorjani, S., et al. Rapid restriction fragment analysis for screening four point mutations of the low-density lipoprotein receptor gene in French Canadians. *Hum. Mutat.* **6:**243–246, 1995.
27. Eiken, H. G., Odland, E., Boman, H., Skjelkvale, L., Engebretsen, L. F., and Apold, J. Application of natural and amplification created restriction sites for the diagnosis of PKU mutations. *Nucleic Acids Res.* **19:**1427–1430, 1991.
28. Friedman, K. J., Highsmith, W. E., Jr., and Silverman, L. M. Detecting multiple cystic fibrosis mutations by polymerase chain reaction-mediated site-directed mutagenesis. *Clin. Chem*. **37:**753–755, 1991.
29. Gasparini, P., Bonizzato, A., Dognini, M., and Pignatti, P. F. Restriction site generating-polymerase chain reaction (RG-PCR) for the probeless detection of hidden genetic variation: application to the study of some common cystic fibrosis mutations. *Mol. Cell Probes* **6:**1–7, 1992.
30. Haliassos, A., Chomel, J. C., Tesson, L., et al. Modification of enzymatically amplified DNA for the detection of point mutations. *Nucleic Acids Res.* **17:**3606, 1989.
31. Lindeman, R., Hu, S. P., Volpato, F., and Trent, R. J. Polymerase chain reaction (PCR) mutagenesis enabling rapid non-radioactive detection of common beta-thalassaemia mutations in Mediterraneans. *Br. J. Haematol.* **78:**100–104, 1991.
32. Newton, C. R., Graham, A., Heptinstall, L. E., et al. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.* **17:**2503–2516, 1989.
33. Sarkar, G., Cassady, J., Bottema, C. D., and Sommer, S. S. Characterization of polymerase chain reaction amplification of specific alleles. *Anal. Biochem.* **186:**64–68, 1990.
34. Wu, D. Y., Ugozzoli, L., Pal, B. K., and Wallace, R. B. Allele-specific enzymatic amplification of beta-globin genomic DNA for diagnosis of sickle cell anemia. *Proc. Natl. Acad. Sci. USA* **86:**2757–2760, 1989.
35. Sommer, S. S., Groszbach, A. R. and Bottema, C. D. PCR amplification of specific alleles (PASA) is a general method for rapidly detecting known single-base changes. *Biotechniques* **12:**82–87, 1992.
36. Ekman, G. C., Billingsly, R., and Hessner, M. J. Rh genotyping: avoiding false-negative and false-positive results among individuals of African ancestry. *Am. J. Hematol.* **69:**34–40, 2002.
37. Ekman, G. C., and Hessner, M. J. Screening of six racial groups for the intron 5 G→A 3′ splice acceptor mutation responsible for the polynesian kidd (a-b-) phenotype: the null mutation is not always associated with the JKB allele. *Transfusion* **40:**888–889, 2000.
38. Hessner, M. J., Luhm, R. A., Pearson, S. L., Endean, D. J., Friedman, K. D., and Montgomery, R. R. Prevalence of prothrombin G20210A, factor V G1691A (Leiden), and methylenetetrahydrofolate reductase (MTHFR) C677T in seven different populations determined by multiplex allele-specific PCR. *Thromb. Haemost.* **81:**733–738, 1999.
39. Hessner, M. J., McFarland, J. G., and Endean, D. J. Genotyping of KEL1 and KEL2 of the human Kell blood group system by the polymerase chain reaction with sequence-specific primers. *Transfusion* **36:**495–499, 1996.
40. Hessner, M. J., Pircon, R. A., Johnson, S. T., and Luhm, R. A. Prenatal genotyping of Jk(a) and Jk(b) of the human Kidd blood group system by allele-specific polymerase chain reaction. *Prenat. Diagn.* **18:**1225–1231, 1998.
41. Hessner, M. J., Pircon, R. A., Johnson, S. T., and Luhm, R. A. Prenatal genotyping of the Duffy blood group system by allele-specific polymerase chain reaction. *Prenat. Diagn.* **19:**41–45, 1999.
42. Hessner, M. J., Shivaram, S. M., Dinauer, D. M., Curtis, B. R., Endean, D. J., and Aster, R. H. Neutrophil antigen (FcgammaRIIIB) SH gene frequencies in six racial groups. *Blood* **93:**1115–1116, 1999.
43. Skogen, B., Bellissimo, D. B., Hessner, M. J., et al. Rapid determination of platelet alloantigen genotypes by polymerase chain reaction using allele-specific primers. *Transfusion* **34:**955–960, 1994.
44. Hessner, M. J., Budish, M. A., and Friedman, K. D. Genotyping of factor V G1691A (Leiden) without the use of PCR by invasive cleavage of oligonucleotide probes. *Clin. Chem.* **46:**1051–1056, 2000.
45. Hessner, M. J., Dinauer, D. M., Kwiatkowski, R., Neri, B., and Raife, T. J. Age-dependent prevalence of vascular disease-associated polymorphisms among 2689 volunteer blood donors. *Clin. Chem.* **47:**1879–1884, 2001.
46. Hessner, M. J., Dinauer, D. M., Luhm, R. A., Endres, J. L., Montgomery, R. R., and Friedman, K. D. Contribution of the glycoprotein Ia 807TT, methylene tetrahydrofolate reductase 677TT and prothrombin 20210GA genotypes to prothrombotic risk among factor V 1691GA (Leiden) carriers. *Br. J. Haematol.* **106:**237–239, 1999.
47. Hessner, M. J., and Luhm, R. A. The C536T transition in the tissue factor pathway inhibitor (TFPI) gene does not contribute to risk of venous thrombosis among carriers of factor V Leiden. *Thromb. Haemost.* **84:**724–725, 2000.
48. Raife, T. J., Lentz, S. R., Atkinson, B. S., Vesely, S. K., and Hessner, M. J. Factor V Leiden: a genetic risk factor for thrombotic microangiopathy in patients with normal von Willebrand factor-cleaving protease activity. *Blood* **99:**437–442, 2002.
49. Aldener, A. and Olerup, O. Characterization of a novel DQB1 (DQB1*0609) allele by PCR amplification with sequence-specific primers (PCR-SSP) and nucleotide sequencing. *Tissue Antigens* **42:**536–538, 1993.
50. Aldener-Cannava, A. and Olerup, O. HLA-DOB1 "low-resolution" typing by PCR amplification with sequence-specific primers (PCR-SSP). *Eur. J. Immunogenet.* **21:**447–55, 1994.
51. Olerup, O. HLA-B27 typing by a group-specific PCR amplification. *Tissue Antigens* **43:**253–256, 1994.

52. Olerup, O., Aldener, A., and Fogdell, A. HLA-DQB1 and -DQA1 typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours. *Tissue Antigens* **41:**119–134, 1993.
53. Olerup, O. and Zetterguist, H. DR "low-resolution" PCR-SSP typing—a correction and an up-date. *Tissue Antigens* **41:**55–6, 1993.
54. Olerup, O. and Zetterquist, H. HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours: an alternative to serological DR typing in clinical practice including donor-recipient matching in cadaveric transplantation. *Tissue Antigens* **39:**225–235, 1992.
55. Wilson, R. C., Wei, J. Q., Cheng, K. C., Mercado, A. B., and New, M. I. Rapid deoxyribonucleic acid analysis by allele-specific polymerase chain reaction for detection of mutations in the steroid 21-hydroxylase gene. *J. Clin. Endocrinol. Metab.* **80:**1635–1640, 1995.
56. Boldt, B., Skogen, B., Agostini, T., Roscetti, D., and McFarland, J. One-tube method for complete HPA-1 genotyping by PCR using sequence-specific primers. *Br. J. Haematol.* **99:**968–973, 1997.
57. Dinauer, D. M., Luhm, R. A., Uzgiris, A. J., Eckels, D. D., and Hessner, M. J. Sequence-based typing of HLA class II DQB1. *Tissue Antigens* **55:**364–368, 2000.
58. Lyamichev, V., Brow, M. A., and Dahlberg, J. E. Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. *Science* **260:**778–783.
59. Lyamichev, V., Brow, M. A., Varvel, V. E., and Dahlberg, J. E. Comparison of the 5′ nuclease activities of taq DNA polymerase and its isolated nuclease domain. *Proc. Natl. Acad. Sci. USA* **96:**6143–6148, 1999.
60. Lyamichev, V., Mast, A., Hall, J., et al. Polymorphism identification and quantitative detection of genomic DNA by invasive clevage of oligonucleotide probes. *Nat. Biotechnol.* **17:**292–296, 1999.
61. Lyamichev, V., Mast, A. L., Hall, J. G., et al. Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nat. Biotechnol.* **17:**292–296, 1999.
62. Lyamichev, V. and Neri, B. Invader assay for SNP genotyping. *Methods Mol. Biol.* **212:**229–240, 2003.
63. Hall, J. G., Eis, P. S., Law, S. M., et al. Sensitive detection of DNA polymorphisms by the serial invasive signal amplification reaction. *Proc. Natl. Acad. Sci. USA* **97:**8272–8277, 2000.
64. Kaiser, M. W., Lyamicheva, N., Ma, W., et al. A comparison of eubacterial and archaeal structure-specific 5'-exonucleases. *J. Biol. Chem.* **274:**21387–21394, 1999.
65. Lyamichev, V. I., Kaiser, M. W., Lyamicheva, N. E., et al. Experimental and theoretical analysis of the invasive signal amplification reaction. *Biochemistry* **39:**9523–9532, 2000.
66. Dawson, E., Chen, Y., Hunt, S., et al. A SNP resource for human chromosome 22: extracting dense clusters of SNPs from the genomic sequence. *Genome Res.* **11:**170–178, 2001.
67. Mein, C. A., Barratt, B. J., Dunn, M. G., et al. Evaluation of single nucleotide polymorphism typing with invader on PCR amplicons and its automation. *Genome Res.* **10:**330–343, 2000.
68. Nagano, M., Yamashita, S., Hirano, K., et al. Two novel missense mutations in the CETP gene in Japanese hyperalphalipoproteinemic subjects: high-throughput assay by Invader assay. *J. Lipid Res.* **43:**1011–1018, 2002.
69. Nevilie, M., Selzer, R., Aizenstein, B., et al. Characterization of cytochrome P450 2D6 alleles using the Invader system. *Biotechniques* (Suppl.) 34–38 and 40–43, 2002.
70. Ohnishi, Y., Tanaka, T., Ozaki, K., Yamada, R., Suzuki, H., and Nakamura, Y. A high-throughput SNP typing system for genome-wide association studies. *J. Hum. Genet.* **46:**471–477, 2001.
71. Olivier, M., Chuang, L. M., Chang, M. S., et al. High-throughput genotyping of single nucleotide polymorphisms using new biplex invader technology. *Nucleic Acids Res.* **30**, e53, 2002.
72. Hessner, M. J., Friedman, K. D., Voelkerding, K. V., et al. Multisite study for genotyping of the factor II (prothrombin) G20210A mutation by the invader assay. *Clin. Chem.* **47:**2048–2050, 2001.
73. Ledford, M., Friedman, K. D., Hessner, M. J., Moehlenkamp, C., Williams, T. M., and Larson, R. S. A multi-site study for detection of the factor V (Leiden) mutation from genomic DNA using a homogeneous invader microtiter plate fluorescence resonance energy transfer (FRET) assay. *J. Mol. Diagn.* **2:**97–104, 2000.
74. Ryan, D., Nuccie, B., and Arvan, D. Non-PCR-dependent detection of the factor V Leiden mutation from genomic DNA using a homogeneous invader microtiter plate assay. *Mol. Diagn.* **4:**135–144, 1999.
75. Tsongalis, G. J., Rainey, B. J., and Hodges, K. A. READIT: a novel technology used in the interrogation of nucleic acid sequences for single-nucleotide polymorphisms. *Exp. Mol. Pathol.* **71:**222–225, 2001.
76. Bean, P., Ryan, A., and Tsongalis, G. J. Point mutation detection technology for the new-century laboratory. *Am. Clin. Lab.* **20:**13–16, 2001.
77. Iovannisci, D. M. The READIT trade mark Assay as a method for genotyping NAT1*10 polymorphisms. *Genet. Test.* **6:**245–253, 2002.
78. Rhodes, R. B., Lewis, K., Shultz, J., et al. Analysis of the factor V Leiden mutation using the READIT Assay. *Mol. Diagn.* **6:**55–61, 2001.

11 Microarray Approaches to Gene Expression Analysis

David Neil Hayes and Matthew Meyerson

1. INTRODUCTION

1.1. BACKGROUND AND OVERVIEW A powerful suite of new technologies now makes it possible to measure simultaneously the expression of many thousands of genes directly from a clinical sample *(1)*. Known alternately as DNA microarrays, gene arrays, expression arrays, "gene chips," or various permutations thereof, technologies for measuring gene expression are already beginning to have many applications in clinical diagnosis. The most promising to date are in cancer diagnosis, where microarray classification has suggested clinically meaningful distinctions that are difficult or impossible to make by conventional histopathology alone *(2,3)*. As the experimental and analytical techniques have developed, applications have been suggested across the spectrum of human pathology, including infectious, vascular, genetic, inflammatory, psychiatric, and metabolic diseases.

Although generally not stated explicitly, the concept of gene expression analysis has been embedded in clinical diagnosis by conventional methods such as flow cytometry, fluorescent *in situ* hybridization (for amplification), immunohistochemistry, and other assays. For example, the characterization of lymphoid malignancies depends on the measurement of surface protein abundance, as detected by labeled antibodies such as the CD group markers *(4)*. Similarly, the presence or absence of estrogen receptor in breast cancer informs key treatment decisions, as does the presence or absence of Her-2/neu amplification *(5)*.

Thus, the new microarray methods represent a quantitative leap in the ability to measure expression of multiple genes and their corresponding proteins, rather than a qualitatively new principle. The power of the microarray lies in its flexibility, quantitative reproducibility, efficiency, and ability to sample the genome in an unbiased fashion. The scope of gene expression arrays in clinical diagnosis is likely to expand dramatically in the near future as increasing numbers of genes and gene products are described with prognostic and treatment significance. In particular, the introduction of molecularly targeted therapies for cancer is likely to drive the demand. Expression arrays are an efficient and robust means of integrating many of these measurements in a single assay.

The goal of this chapter is to discuss the current tools used in functional genomics with an emphasis on clinical applications and focusing mainly on the various microarray platforms. We will discuss issues of study design, sample selection and preparation, and the contribution of measurement error. Whereas genomics analysis is heavily dependent on a variety of computer imaging, software, databases, and bioinformatics tools, we will sample a range of representative applications. We place particular emphasis on the processing of raw data into forms amenable to analysis of underlying biology. The analytic methods used in genomics assays are unlikely to be familiar to most readers and we describe their underlying principals. Acknowledging the technical limitations of the microarray, we will introduce a range of complementary DNA, RNA, and protein-based technologies. In all of these areas, we emphasize that the field of functional genomics is young and few conventions have emerged that apply to all situations the reader might encounter. Despite the ongoing evolution, the basic principles described in this chapter are hopefully sufficient to allow the reader confidence in evaluating current clinical genomics research. To illustrate this point, we identify in a modular way the components of two landmark publications, identifying the relevant methods in light of the lessons we review in this chapter.

1.2. BACKGROUND: GENE EXPRESSION DEFINES DIFFERENTIATED TISSUE The human genome is estimated to contain between 30,000 and 120,000 genes. However, in a given cell or tissue, only a subset of the genes is expressed. The identity of the expressed genes principally defines the character and function of a particular cell type. For example, red blood cells have a very high concentration of hemoglobin, whereas this protein is essentially undetectable in other cells. Accordingly, red blood cell precursors will express high levels of globin genes and other cells will not. Similarly, muscle, heart, and brain are characterized by expression of different isoforms of creatine kinase, a distinction that is useful for the diagnosis of myocardial infarction.

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

Measuring the overall pattern of expressed genes allows the distinction between distinct differentiated tissues. This is particularly valuable in cancer diagnosis, in which accurate diagnosis dictates treatment and diagnosis is often uninformative using current methods. The identification of unique gene expression signatures, associated with different cell differentiation states, permits precise cancer diagnosis. For example, the distinction between acute lymphoid leukemia and acute myeloid leukemia may be made on the basis of global gene expression patterns *(6)*.

2. HISTORICAL PERSPECTIVE

Before entering the mechanics of genomic analysis we will consider three developments that ushered in the current technology. The first event was the refinement of practical genomic assays. These assays are practical in that they are relatively inexpensive, rapid, and widely available. Perhaps the most important was first described in Patrick Brown's lab in the journal *Science* in 1995 *(7)*. Brown and colleagues developed a method in which single-stranded DNA of known sequence could be mechanically spotted onto a glass slide in a grid pattern. Having made the archetype of what is now known as a DNA microarray (or DNA chip), the researchers isolated messenger RNA (mRNA) from samples under a variety of conditions. The mRNA was reverse-transcribed to complementary DNA (cDNA) and fluorescently labeled before being hybridized against the cDNA spotted on their microarry. The resultant hybridized arrays were quantitatively scanned for fluorescence, producing an image in which the intensity of each spot corresponded to the degree of hybridization present for that cDNA clone. Accordingly the intensity of individual spots was proportional to the mRNA present in the initial sample, confirmed by Northern blot. Although refinements have been made since, this method remains the model for the most widely used assay in the field of genomics.

The second defining event in the evolution of functional genomics was a dramatic improvement in the quality of genomics databases. These improvements occurred as a result of the completion of the Human Genome Project, along with the efforts to systematically name genes and curate gene libraries such as Unigene managed by the National Center for Biotechnology Information *(8–11)*. The patterns generated by DNA microarray experiments have potential uses independent of a full understanding of the genes that comprise the arrays. For example, expression sequence tags (ESTs) of known sequence but unknown function are routinely reproducibly associated with specific disease states *(12)*. However, to make full use of the genomic data, researchers and clinicians need to be able to communicate efficiently using such information as sequence and gene function.

The third component in the genesis of functional genomics relates to the technical improvements in microcomputing, imaging, and statistical methodologies. Just as the microcomputer popularized multivariate regression techniques in the 1980s, the 1990s witnessed progress in computing speed and digital imaging necessary to process genomic experiments. Concurrently, biostatisticians, computer programmers, and computational biologist adapted analytical techniques. Finally, dramatic improvements in Internet technologies allowed both data and methods to be shared efficiently.

Each of the three necessary developments continues to evolve. The DNA microarray itself is constantly being refined in tandem with a wide variety of other genomic assays. Similarly, the initial phase of the Human Genome Project has been completed, but the range of normal human DNA variation continues to be defined. Gene annotation databases, such as the Gene Ontology Consortium's, are updated daily as details on function emerge. Finally, the computational standards of genetic analysis are in a constant state of development. In spite of these ongoing developments, a core knowledge base has emerged and will serve as the focus of this chapter.

Although an initial flurry of genomics activity focused on cell culture, proof of concept, methods of analysis, and simple organisms, publications using human samples emerged relatively quickly *(13)*. By 1999, a significant number of clinical samples were being analyzed using array technologies, and by 2000, we saw the first publications in the Abridged Index Medicus journals (a subset of journals devoted to clinical research). An understanding of basic genomics techniques is vital for those who will soon be called on to evaluate the growing number of clinical applications.

The push toward genomics research underscored by a promise that previously insolvent clinical research problems might now have new solutions. Perhaps the most common difficulty is recruitment of patients in sufficient numbers. Recruitment is a problem for rare disease, but even studies of common disease fail to generate cohorts of sufficient size. When the clinical problems are complex or the biology of the system incompletely understood, the theoretical numbers of patients needed could easily exceed all reasonable recruitment expectations. Significant time and effort investment are required to answer a single clinical question, bearing little fruit when recruitment is not met. The losses can be mitigated by testing multiple hypotheses in one study, although the methodology of probabilistic hypothesis testing frowns on this practice.

As we will see in the following sections, clinical genomics offers an alternative to the limitations of clinical research by two mechanisms. First, genomics offers tools for analyzing studies with small numbers of samples, each sample being associated with many variables. Second, genomics analysis offers the potential to clarify complex biologic systems, including identifying disease subgroups. When subgroups are identified with respect to disease behavior, clinical trials could be designed to target those with the largest expected response to therapy. This process could drastically decrease the required sample size needed for such a study.

The problems of patient recruitment, rare diseases, and complex biology are particularly evident in the field of oncology. The ability to identify tumor subtypes that have characteristic biology or clinical behavior would be a major contribution in terms of reduction of disease complexity. There is good evidence that microarray data might refine tumor diagnosis in cases where light microscopy and special stains are indeterminant. Investigators have differentiated carcinomas of unknown primary, primary lung tumors from metastatic colon cancer, and estrogen receptor-positive and -negative tumors *(12,14,15)*.

In addition to identifying know tumors types, previously unrecognized tumor subtypes might be described by clinical outcome, such in the work of Alizadeh in lymphoma or by differences in underlying cancer biology *(16)*. Perhaps for these practical reasons, the promise of genomics techniques in oncology is being embraced the most enthusiastically. An informal Medline review suggests that as many as half of all genomics publications relate to the field of cancer and we will draw on cancer for the clinical examples in this text.

3. DNA MICROARRAY PLATFORMS

In the following subsections, we will develop that DNA microarray model introduced in section 2 by detailing the two most commonly used DNA microarray platforms. The basic principles for the cDNA array remains true to that first described by Schena and Brown. The authors helped ensure wide distribution of the technology by making available protocols for building a robot spotter to produce DNA chips, a scanner to analyze them, and the software to run the experiment *(17)*. With these protocols, a number of early investigators with the technical expertise were able to set up in-house chip production and analysis facilities for $50,000–$100,000 *(18)*. With the initial facility, the marginal cost of producing individual chips was low. Commercially available equipment also appeared in the late 1990s and many institutions adopted the cDNA platform by establishing internal chip production facilities. A number of biotech firms also produced individual chips for sale; however, the individual chip market from the beginning has been dominated by an alternate proprietary technology; the oligonucleotide array, which we will discuss in Sections 3.4–3.7.

The availability of two alternative DNA microarray platforms has led to an ongoing debate over which is superior. We will touch on the strengths and weaknesses of the individual platforms in the following subsections; however, it is premature to declare one technology a clear victor. Instead, the choice of cDNA array vs oligonucleotide array is influenced most by local availability, cost, and institutional experience. Although there are significant differences between the platforms, the underlying design and execution of DNA microarray assays is similar and follows the format shown in Table 1 *(19)*.

Table 1
Elements of DNA Microarray Assay

1. Chip manufacturing: probe selection, chip production
2. Sample preparation: RNA isolation and labeling
3. Chip sample hybridization
4. Scanning of hybridized chip (image acquisition)
5. Image analysis: transduction of spots to genomic information (RNA expression values)
6. RNA expression standardization and normalization
7. Data analysis

3.1. cDNA ARRAY CHIP PRODUCTION

An investigator wanting to use the cDNA array platform can purchase a commercially available chip or more commonly independently manufacture the chip as illustrated in Fig. 1. This review summarizes concepts needed to appreciate the limitations of the technology, noting that protocol details are widely available. Before construction of a cDNA microarray, the investigator needs to select DNA clones, known as "probes" in the vocabulary of the microarray. For the purposes of this review, we will refer to the cDNA (or mRNA) from the sample as the "target," with the spotted DNA termed the "probe." We note that there is not universal consensus on these terms and readers might find them used differently elsewhere. Clone selection is governed by logistic and scientific considerations, of which clone availability is chief. Surprisingly, clone number is a minor component, as 30,000–50,000 probes can be spotted on a single slide, a number that overlaps the estimated total number of human genes. In some cases, researchers use relatively small numbers of clones of particular interest or that are readily available. More commonly, clones are purchased in bulk from a commercial supplier such as the 9128 Human Unigene 1 clone set from Incyte. The quality of these sets can vary significantly and it is usually of interest that the source be identified. In their usable form, the clones are single-stranded DNA of 0.6–5 kb representing entire genes derived from cDNA, partially sequenced genes, or ESTs.

Once selected, the DNA itself needs to be prepared for spotting onto a solid medium, usually a coated glass slide, in a manner represented in Fig. 1. Approximately 100–500 ng of DNA are isolated and purified from bacterial vectors and placed into wells in preparation for transfer to the slide. The transfer from well to a preselected location on the glass slide is via a robot spotter using a variety of spotting technologies ranging from a fountain pen pipet mechanism, to capillary tube, to noncontact ink jet printer *(18,20)*. Each transfer modality was designed to overcome the challenge of working with minute liquid volumes while obtaining spots of consistent size, 100–200 μm. Each spot, even at this size, is comprised of millions of copies of the DNA sequence of interest.

Prior to the spotting process, the surface of the glass slide is coated with an organic compound such as silane to permit stable DNA binding. Probes might make multiple contacts with the organic matrix along their lengths, with longer sequences interacting the most. Multiple surface interactions could decrease target DNA binding proportional to sequence length, not mRNA concentration, a potential probe quality concern. Other local phenomena such as variable drying could also interfere with eventual DNA hybridization *(21)*.

In addition to technical challenges of generating high-quality clones and spotting them accurately, there are critical data management challenges in microarray design *(22)*. For example, a clone might represent only a segment of a gene or one of several alternate splices; therefore, it is often insufficient to identify a spot by the genes name alone. Furthermore, many clones are ESTs, without consensus gene name. To explicitly identify a spot, the investigator needs to link the probe sequence both to its location on the slide and the well from which the clone was selected. Any break in the chain of information can make interpretation of the assay impossible.

3.2. cDNA ARRAY SAMPLE PREPARATION AND HYBRIDIZATION

To prepare RNA for use in a cDNA array, total RNA is isolated from a sample of interest by one of the standard methods *(23)*. Ususlly, 50–100 μg of total RNA (2 μg of mRNA) is sufficient, although this could vary by protocol.

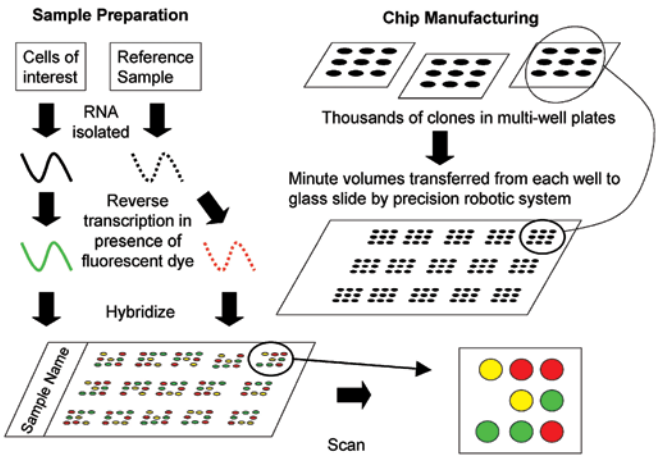


Fig. 1. cDNA microarray manufacturing process. See text for description. (Courtesy of the NCBI Gene Expression Omnibus.) (Figure appears in color in insert following p. 172.)

Because of the rapid degradation of RNA, only fresh or fresh frozen samples generally provide RNA of sufficient quality for analysis. Even samples that have been collected for the specific purposes of RNA analysis should be evaluated for overall RNA quality.

Most contemporary cDNA array experiments require the simultaneous analysis of RNA from two separate sources on the same chip: an experimental biologic specimen and a reference. In separate reactions for experimental and reference, total RNA is reverse-transcribed to cDNA using either random primers or an oligo-dT primer. The oligo-dT primer takes advantage of the presence of a poly-A tail of mRNA to selectively transcribe only mRNA species. Reverse transcription of the experimental and reference RNA is carried out in the presence of different fluorescent labels, most commonly cyanine 3 (Cy3), which fluoresces green, or cyanine 5 (Cy5), which fluoresces red. Once labeled, experimental and reference cDNAs are combined in solution and poured over the chip to allow hybridization, with experimental and reference targets competing for the same probes. After a prescribed hybridization time, the slide is dried and prepared for scanning.

3.3 cDNA ARRAY IMAGE ACQUISITION AND ANALYSIS

A variety of optical systems have been developed for detecting fluorescence, of which we will consider only one *(24,25)*. A laser tuned to the wavelength of the Cy3 dye is passed over the slide while a photodetector quantifies green fluorescence at a resolution of approx 5–20 μm. In this manner, a 16-bit tagged image format (TIF) image is created in which the intensity of each green spot is proportional to the quantity of labeled cDNA found at a given location on the array. The process is repeated with a laser calibrated to the Cy5 dye fluorescence. The two color channels, one green and one red, are superimposed to create a single image such as that shown in Fig. 2. Note that the image used in Fig. 2 was selected to include a high percentage of artifacts that will be discussed here and grossly understates typical image quality. In the two-channel image, the relative amounts of red and green correspond to relative amounts of mRNA from the experimental and reference specimens. When experimental and reference groups are labeled with Cy3 and Cy5, respectively, a spot that is pure green represents mRNA expressed only in the experimental sample. Similarly, a pure red spot corresponds to an mRNA species found only in the reference sample. If neither sample expresses any mRNA, the spot will be black. If mRNA is expressed both in experimental and reference samples, the spot will appear as shades of yellow, orange, or brown.

Figure 2 illustrates the heterogeneity of spots shapes in addition to variation in color and intensity already described. In some cases, spot size varies with signal intensity, with more intense signals having a saturation effect similar to that produced when photographing a bright light. In other cases, however, even dim spots appear quite irregular. Irregularities are most commonly the result of subtle mechanical effects encountered in the original chip spotting process. Small changes in the quantity of DNA spotted, the pressure applied to the slide, or variation in the individual pins used in spotting cause variation in spot size. Variations in fluorescence for a given spot are not only proportional to the amount of mRNA in the sample of interest but they are also proportional to spot topography. To account for variation in spot size in the analysis, the concept of a reference sample was introduced. In place of total fluorescence at each spot, the ratio of fluorescence between an experimental and a reference sample is calculated, a value independent of variations in spot size and shape.

Although ratios overcome the problem of irregular spots, they introduce a new set of challenges *(26)*. The first is that ratios can be unstable measures. When RNA from one of the samples comprising the ratio is present in small quantities

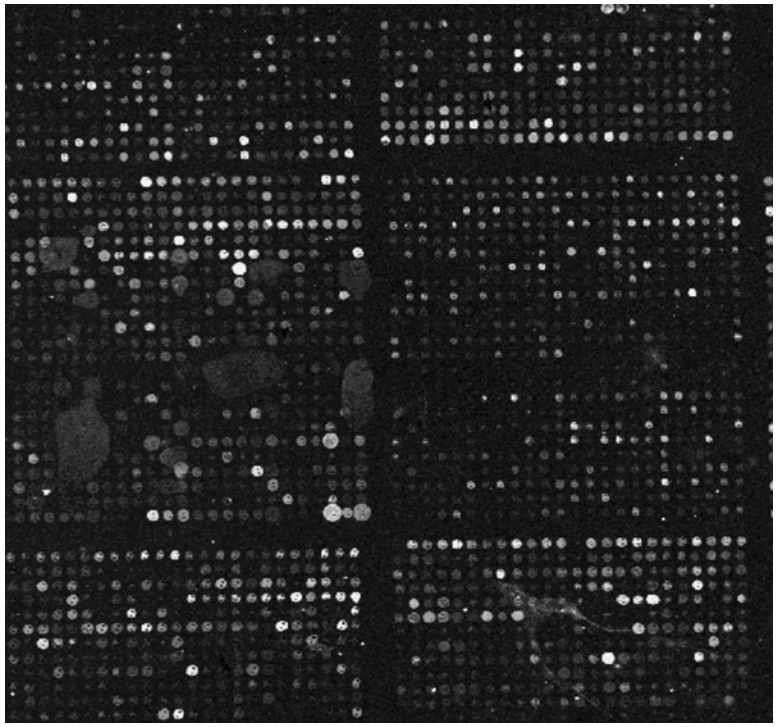


Fig. 2. Example of cDNA microarray image file. Magnified view showing approx 10% of the "spots" from a selected cDNA microarray. *Note*: This image was selected for its unusually large number of artifacts, including smears and streaks. It is not meant to represent the quality of a typical array. See text for details. (Figure appears in color in insert following p. 172.)

relative to the other, then the ratio will be unstable. For example, if the experimental sample has a fluorescence of 10 units and the reference has fluorescence of 0.2 units, then the ratio of experimental to reference fluorescence is 50. A 1% relative error in the reference channel could have resulted in measures of 0.1–0.3, resulting in ratios of 33–100. A second concern in using ratios is the failure to differentiate between expression changes that occur on a small or large absolute scale. For example, a ratio of 0.1 does not differentiate between the comparison of 1/10 and 100/1000. Although not all challenges in the interpretation of ratios have been addressed, many have been accommodated through analytic tools and study design.

The problem of the unstable ratio has many examples in cancer biology, as well as a potential solution through careful study design. The comparison of a tumor expressing an oncogene at very high levels to normal tissue that does not would be potentially unstable. One solution, applied in the study design phase, is to avoid the use of "normal" tissue as the reference for the ratio. Investigators have opted to use cell lines or groups of cell lines that express a wide variety of genes to avoid the problem of low expression in the reference sample. Cell lines also have the theoretical advantage of offering a reproducible reference, whereas "normal" human tissue might be quite variable depending on the source *(27)*.

Using "normal" RNA as the reference has the advantage of making the comparison of experimental to control groups on the same slide. However, as the previous example suggests, the reference sample is not necessarily the experimental control, but rather a means of correcting for variable spot size. Although the direct comparison is initially appealing, there are theoretical considerations beyond those of estimating stable fluorescence ratios. First, by measuring the same normal control on every slide, the investigator has many measurements of normal, yet only single measurements of individual disease samples. Repeated measures of normal in this way might represent an inefficient use of data. Second, fluorescence ratios include two error components: one from both the experimental and the reference sample. In an analysis comparing disease to normal, the two error components are a normal part of study interpretation. When contrasting two disease states (each using normal as a reference), the error introduced by the ratio represents unnecessary noise in the system. Study designs to increase efficiency and address these concerns have been developed, although they are admittedly more complicated to interpret. One such example, a series loop experiment uses the reference RNA for one chip as the experimental RNA on a second chip *(28)*. Measurement error and optimal reference RNA are considered further in later sections and we refer the interested reader to full reviews *(29)*.

In addition to the grid of green, red, yellow, and blank spots, most cDNA image files demonstrate a variety of irregular streaks and spots. Figure. 2 was selected explicitly to illustrate an extreme range of such defects, noting that a more typical field of view would have far fewer artifacts or none at all. Here, we see irregularities resulting from background noise from sources such as dust, local drying effects, and mechanical spotting difficulties. Much of this noise can be attenuated through software and analytical techniques involved in image processing, but arrays should generally be inspected for severe artifacts. Commercial arrays are often shipped with quality control measures in place to minimize these concerns.

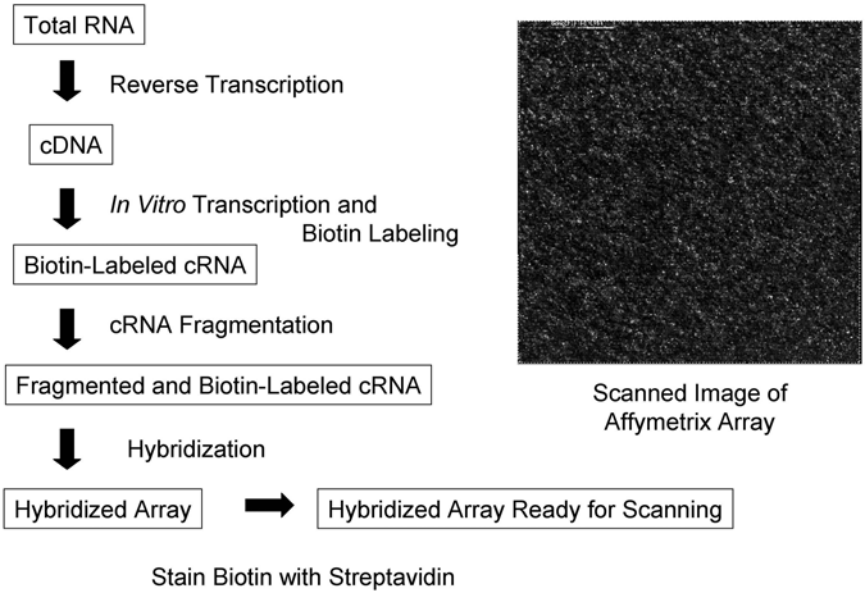


Fig. 3. Summary of Affymetrix oligonucleotide microarray protocol.

Once the two-color image has been generated, the final step in preprocessing the array is to calculate fluorescence and ratio values for each of the spots and associate those values with the genes that they are intended to represent. Calculation of a summary fluorescence statistic is more involved than might be apparent at first glance, as the spots are highly variable over their areas. There might be a dim focus in the center, a gradient of intensity, or other complex topography that needs to be reduced to a single value of fluorescence for both the red and green channels. The public-domain software ScanAlyze written by Michael Eisen has been widely used for this purpose, but other programs are available. Finally, the data are exported to a spreadsheet that links the gene identities of the spots with the fluorescence ratios as well as values of each channel.

3.4. OLIGONUCLEOTIDE ARRAYS: AFFYMETRIX PLATFORM AND CHIP PRODUCTION In parallel with the development of the cDNA microarray, an alternate proprietary technology called the oligonucleotide microarray was made available (Figs. 3 and 4) *(30,31)*. Unlike cDNA arrays, the chips themselves cannot be manufactured by individuals; they must be purchased from the manufacturer (Affymetrix of Santa Clara, CA). Although the oligonucleotide arrays share all of the features of a microarray assay described in Table 1, the differences between the platforms influence study design and data interpretation.

Two technological advances paved the way for the manufacture of oligonucleotide arrays: photolithography and confocal fluorescence scanning. Photolithography as illustrated in Fig. 5 is a process in which oligonucleotides are synthesized directly onto a solid matrix. The process starts with the substrate—a quartz wafer chosen for its optical properties and hydroxylated surface that acts as a linker to which the oligonucleotides are attached. The quartz surface is coated in silane as in the cDNA example and then a synthetic linker with a photochemicallly removable group is attached. A chrome mask with a grid pattern of apertures is precisely aligned over the chip. Ultraviolet radiation is applied to the system such that only the areas directly under the apertures of the mask, an area of 18–20 μm, are irradiated. The photosensitive group is removed from the linker in the irradiated sectors only, unveiling a binding site for a single nucleotide. The surface of the chip is then bathed in a single-nucleotide species (adenine, thymine, guanine, or cytosine) to which an additional photochemically removable group is attached. A new chrome mask can now be applied and the entire process repeated. In this manner, short oligonucleotides (usually 25 nucleotides long [called "25-mers"]) can be constructed. The sequence of the nucleotides is dictated by the order in which the masks and nucleotides are applied. The individual chips are 1.28 cm^2 and capable of harboring over 500,000 unique oligonuclleotide locations. Again, as with the cDNA arrays, each location on the array contains millions of copies of the unique oligonucleotide.

Whereas the cDNA microarray spots are composed of DNA sequences of 600–5000 bases, the oligonucleotide sequences of 25 bp seem quite short. The short sequence of uniform length addresses the concern (raised in Section 3.2) for varying DNA hybridization by strand length, but it raises concerns for lack of specificity. The oligonucleotide microarrays improve specificity by combining groups of 25-mer probes to form a "probe set." It is the probe set, of which each probe is only a part, that is specific for a gene target. Probes in a set are selected from a variety of transcribed regions along the gene using software algorithms and empiric testing, which are best described in the technical notes provided by Affymetrix. Figure 6 is an example of the distribution of probes along the length of a sample probe set. Probe sets vary in the number of probes they contain depending on the chip. The U95 human chips, for example, used probe sets consisting of 16 probes, representing 400 bp along the length of the target gene, a size sufficiently large to convey specificity. For the sake of this review, we will discuss probe sets as they were constructed for the U95 GeneChip® array.

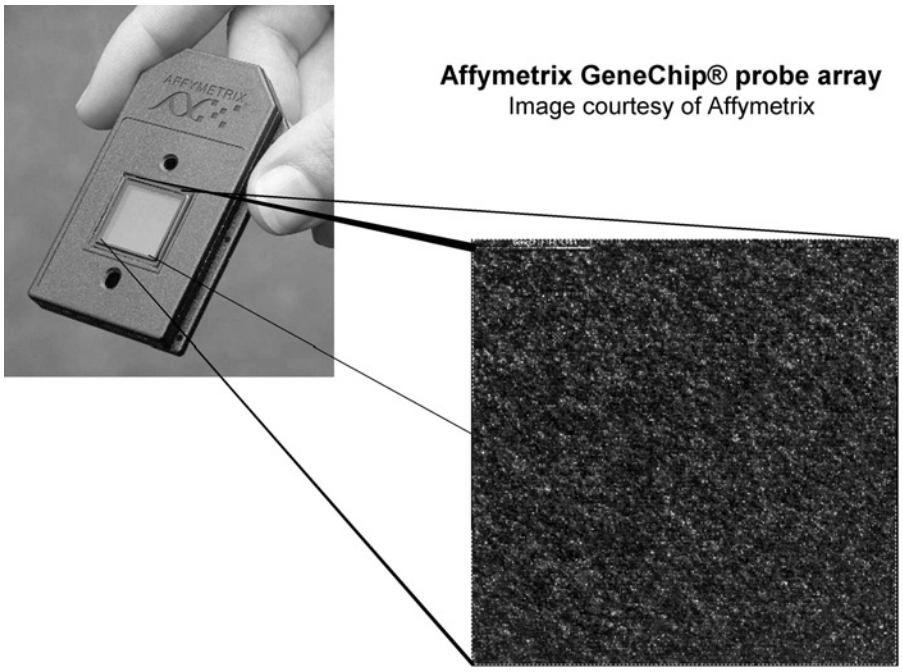


Fig. 4. Example of Affymetrix GeneChip array and microarray image file. (Courtesy of Affymetrix.) (Figure appears in color in CD ROM.)

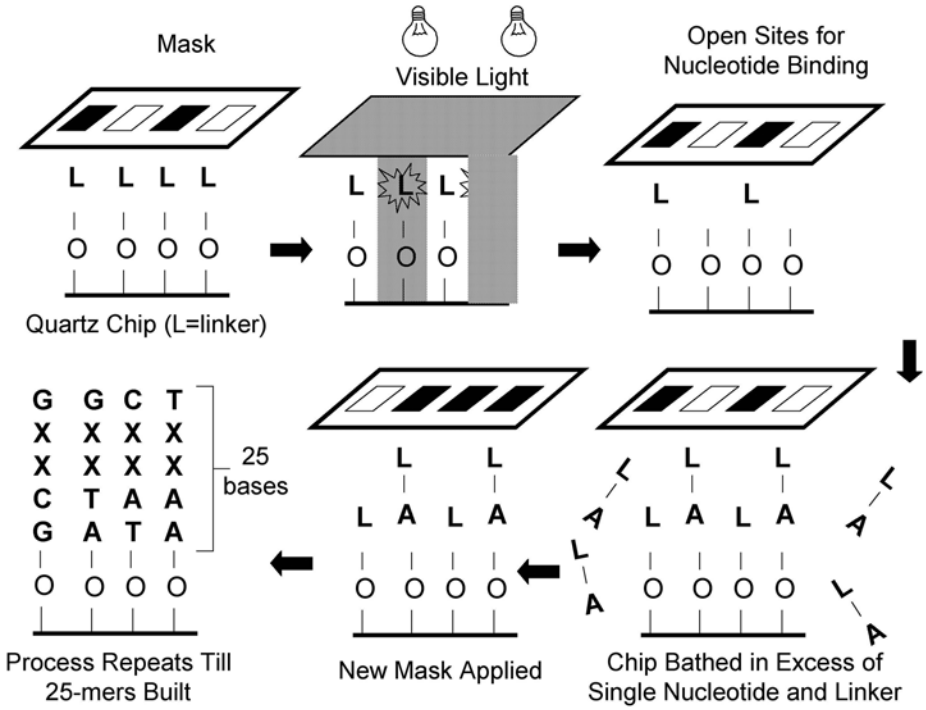


Fig. 5. Photolithography as applied to Affymetrix GeneChip manufacturing.

A major difference between the oligonucleotide array and the cDNA array is that only one RNA sample is hybridized per array in the oligonucleotide platform. The consistency of the photolithographic process obviates the need to control for spot size. Background noise, as well as nonspecific binding to any individual 25-mer necessitate the inclusion of an additional feature on the Affymetrix chips. For each of the 16 probes in the probe set, there is a second probe called a mismatch (with the original probe being called the "perfect match"). The mismatch probe has exactly the same sequence along its 25-bp length except for a single substitution at position 13. Hybridization to the mismatch probe is used to calibrate the perfect match, accounting for nonspecific binding and background noise.

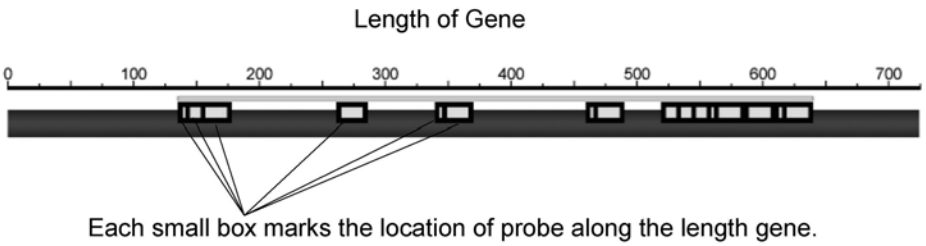


Fig. 6. Example of Affymetrix probe distribution along a gene. Each small box marks the location of probe along the length gene. There are 16 boxes corresponding to the 16 probes in a probe set. (Figure appears in color in CD ROM.)

3.5. OLIGONUCLEOTIDE ARRAYS: AFFYMETRIX PLATFORM: SAMPLE PREPARATION AND HYBRIDIZATION Sample preparation varies somewhat from the protocol used in cDNA array analysis as seen in Fig. 3. The starting material is again total RNA, although 5–10 μg of high-quality RNA is sufficient. Again, the total RNA is reverse-transcribed using the oligo-dT primer, although no fluorescent label is added in this step. The cDNA is in vitro-transcribed to cRNA in the presence of biotinylated deoxynucleotide triphosphate, which will serve to label the RNA with streptavidin-conjugated fluorescent markers after hybridization. In a final preparatory step, the cRNA is fragmented and then hybridized against the oligonucteotide array.

3.6. OLIGONUCLEOTIDE ARRAYS: IMAGE ACQUISITION AND ANALYSIS The chip is then washed and the hybridized cRNA labeled via the biotin–strepavidin system. Fluorescence proportional to the degree of hybridization is recorded by confocal scanning, in which a laser induces fluorescence on one side of the chip while an optical scanner records signal intensity on the other. In this way, a single channel black-and-white image is generated. Each spot represents an oligonucleotide probe, the intensity which is, in theory, proportional to the concentration on mRNA present in the original biologic sample. The raw image produced in this way is saved as a .dat image file, with information on pixel intensity measured similarly to the .tiff file generated in the cDNA example (Fig. 4).

Unlike the cDNA array, the oligonucleotide array does not require the intensity of each spot be interpreted in relation to a reference. Intensity can be interpreted directly as a linear unitless measure called "expression." However, because a given gene target is represented by a probe set of perfect matches and mismatches, the gene expression value needs to be computed from data provided by the entire set of 32 probes. This process begins at the image processing stage when probe-level expression is calculated by software such as the proprietary Microarray Suite from Affymetrix or the public-domain program dChip *(32)*. This step is analogous to that performed on cDNA arrays by Scanalyze, however, unlike Scanalyze, the output is not a table of gene fluorescence values or ratios; it is an intermediate file with the file extension. CEL with probe-level data only. Both Microarray Suite and dChip will integrate the perfect matches and mismatches by a variety of additive and averaging methods such as the AvDiff employed by Microarray Suite v4.0. Using the AvDiff method, the difference in intensities of the perfect match–mismatch pairs are summed and averaged to calculate a final summary statistic "gene expression" for each gene on the array.

$$\text{AvDiff} = (1/J)_1 \Sigma^j (\text{Perfect match}_j - \text{Mismatch}_j)$$

where J is the number of suitable probe sets. There is no consensus standard for integrating across probes in a probe set, with several reasonable alternatives currently in use *(33–35)*. The final output of the assay is a data table with expression value for each of the genes on the array. Because the oligonucleotide arrays are a commercial product in which the probe locations have been standardized and incorporated into the software, the burdens of data management described in the cDNA example are clearly simplified.

Further data standardization and normalization described in a later section are routinely required prior to analysis. For those planning to use oligonucleotide array data generated by other investigators, it is important to either obtain the raw image files or have a full understanding of how gene expression for the data were calculated. Similar warnings apply to the use of cDNA ratio values, which vary depending on the software used to interpret the image file.

3.7. OLIGONUCLEOTIDE ARRAYS: AGILENT PLATFORM The chrome mask used by Affymetrix is limiting in that the design of new chips requires that new masks first be manufactured. To the extent that building new masks is expensive and slow, the system is relatively inflexibility for probe optimization or new array customization. Alternate approaches technologies to produce oligonucleotide arrays have been developed *(36–39)*. Of these the most widely available is the inkjet method developed by Agilent Technologies in which oligonucleotides are printed onto the surface of a glass slide. Essentially, the process mimics that of commercial ink jet printers, with the four ink colors replaced by the nucleotides ATCG.

In place of the photoprotective site used in the Affymetrix platform, a covalently bound trityl group blocks the 5′ hydroxyl group of the nucleotide. As with the other array platforms, the process starts with a silane-coated glass slide with hydroxyl groups available to initiate the oligonucleotide construction at specific loci. The printer head scans the surface of the slide, delivering 100 pL volumes of single-nucleotide species. Unlike the Affymetrix platform, all four nucleotides can be deposited on the slide in a single pass—one base per spot on the array. Nucleotides are deposited in excess of available binding sites, forming 100-μm-diameter spots with 30-μm intervening spaces. Once bound to the glass, excess nucleotide is washed away and the trityl moiety is removed from the 5′ end of all

oligonucleotides. With the trityl group removed, the 5' hydroxyl group is available for strand elongation when the next nucleotide is deposited from the printer system. The combination of precision electronics, surface tension, and orientation of linker molecules ensures that microdroplets are precisely delivered and that oligonucleotide construction proceeds efficiently.

The advantage of this system over that used by Affymetrix is that changes in the chip design require only changes in the software running the printer array, not the construction of a new series of masks. Underperforming probes can be refined with relative ease, new probes tested, and the entire system iteratively optimized at a relatively discounted cost. Product development of this type have lead Agilent to favor gene measurement using oligonucleotides of 60 bases in length and not the probe set methodology employed by Affymetrix. Arrays produced in this manner appear to perform comparably to the other platforms, although investigators generally have less experience with them.

4. SAMPLE SELECTION

In addition to the RNA isolation techniques discussed in Section 3.2, there are a number of sample selection issues key to the proper interpretation of microarray assays *(27)*. Sample selection includes considerations of the population of patients selected for the study, the selection of tissue within the patient, and method of selecting cells for RNA collection. The proper selection of patients cannot be overemphasized because no analytical plan can overcome a major error in sample selection. For example, a microarray analysis of two groups of patients—one sick and one well, in which all sick patients are female and all well patients are male—will identify sex-specific genes as being associated with disease. This extreme example reminds us that the epidemiologic principles of confounding are still present even in genomics experiments. A more subtle but related observation is the following. Microarray assays are sensitive to experimental conditions, and arrays performed at different facilities or under different conditions will show systematic variations called batch effects. If all samples with a given phenotype are performed in one lab and all samples with another phenotype are performed at a second lab, it can be difficult to isolate batch effects from biologic effects. Batch effects can be overcome by ensuring that samples are analyzed in a consistent manner and careful study design.

The earliest microarray studies were performed on uniform populations of cells such as cell culture or the so-called liquid tumors of leukemia and lymphoma. Uniform samples such as these are not representative of the majority clinical relevant specimens. Tumors are composed of varying percentages of inflammatory, stromal, normal, and malignant cells. Each of these cell populations, including the cancer cells, might be comprised of a variety of subpopulations. At the gross level, the investigator needs to assess which portions of a sample are appropriate for analysis. Selection of an obviously tumorous section of a specimen might include necrotic areas of poor RNA quality. Sectioning at the edge of a tumor might include a large percentage of normal cells. Some tumors, such as Hodgkin's disease and prostate cancer, frequently have large amounts of nonmalignant cells throughout the tumor. There is no consensus on which approach is correct, but the investigator should at least consider how selection might influence the study outcome. Early concerns on the usefulness of microarrays have focused considerable attention on proper sampling with these issues in mind.

4.1. SPECIAL TECHNIQUES One approach to microscopic sampling has been to estimate percent tumor by light microscopy and by sample RNA from tissue sections immediately adjacent. More explicit techniques to sample only cells of interest from a heterogeneous background have been developed, including cell culture, flow cytometry, needle dissection, and, most commonly, laser capture microdissection (LCM). In LCM, a tissue section, either frozen or paraffin embedded, is placed on a glass slide and covered with a thin film of ethylene vinyl acetate. Under direct microscope visualization, a low-power infrared laser is directed at areas of interest, heating the polymer above the cells. After laser capture in this manner, the cells adherent to the film are removed for analysis *(40)*. The resolution of the methods allows for capture of single cells without degradation of the RNA.

Laser capture microdissection has been used by many investigators for genomic profiling of tumor cells *(41–43)*. In other cases, however, investigators have argued that tumors are complex systems in which the microenvironment of interaction between cells is critical. Investigators have shown that tumors with a low percentage of malignant cells can have distinct profiles, such as the clear differentiation of colon metastases from lung cancer in samples of 30–50% malignant cells. The debate regarding proper sample preparation has not yet been resolved. Most likely, the two approaches each have different merits, depending on the question the investigator is trying to answer *(44,45)*. If the concern is to clarify an aspect of tumor cell biology, perhaps selective sampling is appropriate. If the interest is gross tumor behavior, then whole tumor sampling is preferred.

Selective sampling methods such as LCM and some nonselective processes such as fine-needle aspiration (FNA) biopsies will often fail to collect sufficient RNA for microarray analysis. A single LCM session will typically select 10^3–10^4 cells, at most a few nanograms of RNA, far short of that required. Similarly, many clinically available samples including but not limited to FNA also harvest insufficient RNA. To overcome this shortcoming, a variety of RNA amplification techniques have been developed *(46–49)*. Initial validation in a variety of clinical and experimental settings demonstrate that RNA-amplified specimens will likely be acceptable for microarray analysis *(41–43)*.

4.2. SELECTION OF NORMAL Mentioned in Section 3.2 were technical concerns for selection of a reference sample for cDNA microarrays that relate to the problem of variable spot size. The goal of reference selection was to achieve a nonzero-level mRNA expression across all genes of interest so that stable ratios could be calculated. Many investigators have used "normal" RNA for that purpose, although, as we will discuss, the selection of control group involves different concerns than the selection of the reference RNA. The selection of a control group, be it normal tissue or otherwise, applies equally to olgonucleotide and cDNA arrays.

In considering what constitutes a normal control, the investigator can either collect histologically normal tissue from the

same diseased individual or normal tissue from another source. Uninvolved adjacent tissue is often used for this purpose, although there is ample evidence that such tissue might be influenced by local molecular effects related to a disease process such as cancer. These histologically normal but molecularly abnormal controls might mask underlying biology of interest in the analysis. Alternatively, the use of tissue from another individual might reveal gene expression differences that do not relate to the disease of interest. Again, consider the comparison of sick females to healthy males in which sex-specific genes confound the relationship between disease and normal. The usual confounders of age, ethnicity, medications, and others still need consideration in the study design and interpretation. Likewise, the diseased group might be heterogeneous in factors unrelated to the biology of their illness that could influence gene expression and deserve similar attention.

5. MICROARRAY STUDY DESIGN AND SOURCES OF ERROR

We briefly mentioned measurement error in Section 3.2 with regard to fluorescence ratios; however, measurement error enters into the array experiment at multiple steps *(29)*. Consider the following experiment for potential sources of error. Liver biopsies are performed on two different mouse populations—one with normal livers and the other with tumors. Messenger RNA is prepared from each of the biopsies. The RNA is labeled and then hybridized against a microarray, and the array is scanned. The use of replicate samples has provided considerable insight into the contributions of measurement error at each of the steps and techniques to minimize that error *(50)*.

There are two general categories of replicate samples: biologic and technical. A technical replicate is the replication of any part of sample preparation or analysis occurring after the biopsy. Starting at a distal aspect of the experiment, we could replicate each probe on the array to measure the reproducibility of fluorescence between probes on the same array. These experiments have been done and show that the agreement of probes on at array is very acceptable at 95–99% *(29)*. Therefore, from the measurement error perspective, there is little value in having replicate probes on an array. Moving proximally in the experiment, investigators have examined the effect of splitting an aliquot of labeled cDNA or, even more proximally, preparing two separate mRNA samples from the same tissue sample before hybridizing to two different cDNA arrays. In these examples, the correlations between genes drops to 80% and 60%, respectively.

The statistics on reproducibility are important in that they describe both the limitations of the technology and suggest methods for improving overall reproducibility. Replicate experiments reveal that technical factors accounts for significant variation in the results. They also suggest the means for reducing this variation considerably. By performing either replicate arrays for the same sample and averaging the results or (more economically) pooling separately labeled specimens, the results will tend toward a more reproducible mean *(50)*. Although technical replicates are not required to obtaining meaningful results, they can be used to reduce error if needed.

Biologic replicates are either repeat biopsies of the same mice or biopsies of different mice with similar tumors. Gene correlation can fall to 30% or lower when biologic replicates are considered. The 30% value includes all of the variations described by the technical components of sample processing and additional variation in the underlying biologic system. Although a reproducibility of 30% might seem low, we should recall that the majority of the genes being sampled likely have little to do with the biologic pathways of interest. A 30% correlation does not mean that each gene is 30% correlated, but, rather, that some are highly correlated and some not at all. High correlation would be expected for the small number of genes with direct relevance to the disease process of interest, with weaker correlations for genes indirectly related to that metabolic state. More exhaustive reviews of DNA microarray reliabilities are available *(28,50,51)*.

6. RNA VS DNA OR PROTEIN

Having developed expression arrays as a model genomics tool, it is worth focused discussion of the implications of using RNA as opposed to DNA or protein. The measurement of RNA is associated with a number of challenges, the most considerable of which is that it is rapidly degraded. The degradation is so rapid that high-quality RNA can only be obtained when tissues is processed immediately or fresh frozen within minutes of collection. Not only is RNA difficult to obtain, as Brown and Botstein have suggested, protein—specifically protein activity—is the biologically active species, not RNA *(52)*. Unlike RNA, many proteins are relatively stable and can be found in paraffin-embedded tissues. As RNA is measured in genomic terms, proteins and protein activities can be measured en masse via proteomic techniques that will be described in Section 8. The problem for current investigators is that the field of proteomics is far less advanced than expression genomics. Proteomics measurements are more complicated, less well standardized, and the researchers less experienced.

The preference for using RNA is not merely convenience. The quantitative relationship between transcription and translation has been confirmed in many settings, including microarrays, arguing for the use of RNA expression profiles *(53,54)*. For the measurement of any single mRNA species, however, other methods such as Northern blot are superior to microarrays *(55)*. Similarly, the usual limitations of RNA measurement remain true in that correlations between RNA and protein levels are low in cases where proteins are secreted, rapidly degraded, or unusually stable. Cellular processes that are decoupled from translation, such as responses to micronutrients, drugs, or physiologic conditions might be poorly represented by changes in RNA levels. Finally, RNA, like protein, is not uniformly degraded and relative changes in RNA species might be artifacts of degredation.

7. OTHER GENOMIC ASSAYS

Expression microarrays are the most widely used genomic assays; however, there are a wide variety of genomic and proteomic tools using RNA, DNA, and protein. Many of these techniques address specific shortcomings of the expression arrays or offer complementary information *(56)*.

7.1. SAGE A distinct weakness of expression microarrays is the use of fluorescence, an indirect measure of mRNA

concentration. We have already discussed the challenges of using the fluorescence ratio as a measure of RNA in the cDNA microarray. In a following section, we will encounter similar standardization difficulties in the oligonucleotide platform. Serial analysis of gene expression (SAGE) is an RNA genomic technique that predates the expression array, in which mRNA copy number is quantified directly, overcoming a distinct shortcoming of the DNA microarray *(57)*. The SAGE technique works on the principle that short oligonucleotides of 10 bases in length, called tags, are sufficient to uniquely identify any specific cDNA transcript. The principle is a statistical assertion based on the approximation that the total number of human genes is around 30,000–40,000 with 80,000–120,000 transcripts based on alternate splicing. A 10-bp tag, comprised of the 4 bases ATCG, has a maximum of 4^{10} possible random combinations, a number far greater than the estimated number of human transcripts. By this logic, it is statistically unlikely that any single 10-bp tag would present more than once. Although we know that DNA sequences are not random and that related genes might often share sequence homology, such events are rare enough to allow SAGE to be useful.

Messenger RNA is isolated from the cells of interest and double-stranded (ds) DNA is produced by reverse transcription, as seen in Fig. 7. The ds-cDNA is immobilized on streptavidin-coated magnetic beads and the cDNA cleaved with a restriction enzyme that cuts most transcripts at least once. Only the 3′ end of the cDNA is retained by collecting the magnetic beads. By isolating the DNA fragment closest to the 3′ end which is cut by a specific restriction enzyme, the 10-bp tag is standardized. In addition, the availability of genomic sequencing allows for the identification of the specific gene from the 10-bp tag. In the example shown in Fig. 7, the restriction enzyme recognizes the site CATG. For this reason, the sequence CATG will be included in all of the subsequent tags in the analysis, although it adds nothing to the specificity of the tag. It is therefore technically correct to say that the tags are 10–14 bp long, of which 10 bp are unique.

A linker is added to the cut end, and a second restriction enzyme, called a type II restriction enzyme, is added. The type II restriction enzyme has the desired property of cutting a DNA sequence at a specific number of basepairs downstream from the recognition site. In this case, the cut produces a 10-bp sequence, the above-described tag, from original cDNA after accounting for the linker. The individual tags are randomly blunt-end ligated to form dimers called di-tags. Di-tags are PCR amplified in a manner to maintain the proportional representation of each tag's frequency in the original mRNA sample. After amplification, the linkers are removed and the di-tags are allowed to concatimerize, a process in which di-tags form long chains. The concatimerized di-tags are subcloned into a vector and their DNA is sequenced. The resultant sequence information is called a SAGE library and can be analyzed to identify and quantify each of the 10-bp tags.

Unlike expression arrays, there is a direct quantification of each of the mRNA species that was present in the original specimen, expressed as the number of copies per 10,000 tags. Whereas expression array data require considerable effort to compare one study to another (a point we will return to later),

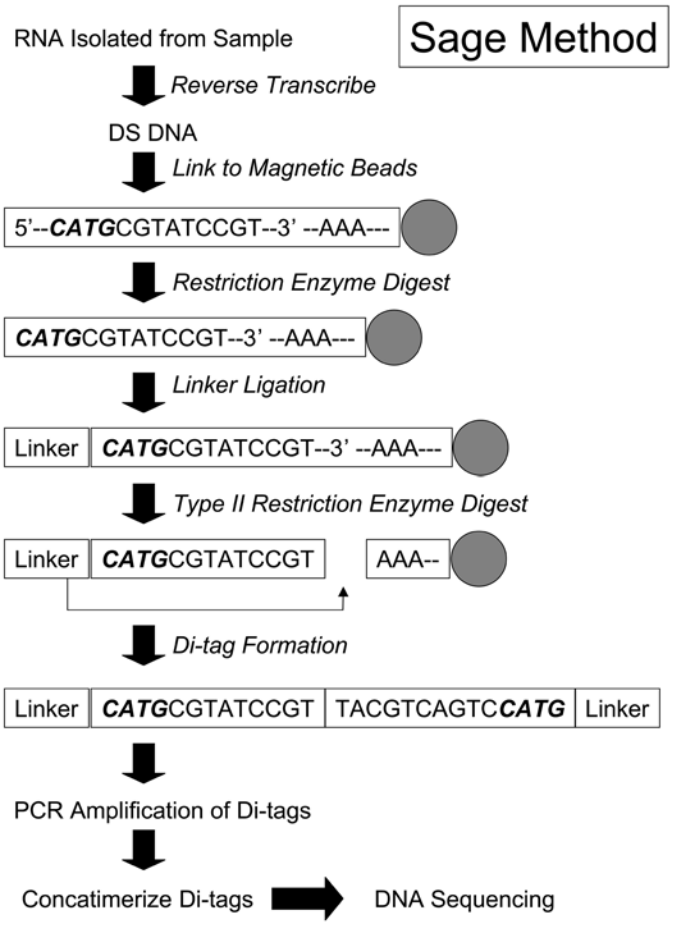


Fig. 7. SAGE method. (See text for details.)

tag counts from different experiments are directly comparable. Researchers from different labs can exchange SAGE libraries and compare experimental results *(58)*. The SAGE method is often used in situations in which expression arrays could also be considered, comparing the expression patterns between tumor and normal cells *(59)*. In fact, SAGE along with the Northern blot has been used as a validation of the microarray technology *(60)*. SAGE offers some advantages over microarrays in the discovery of novel genes, however. Although expression arrays usually require that a DNA sequence be known to be included on the array, SAGE samples all transcribed DNA. Limiting its broad use, however, is that compared to microarrays, the preparation of a SAGE library is laborious and expensive.

7.2. DNA-BASED GENOMIC ASSAYS: CGH, ARRAY CGH, AND SNP ARRAYS Thus, far we have used RNA quantity, as measured by RNA expression or cDNA fluorescence, to describe the activity of genes across the genome. There are also techniques reliant on genomic DNA to measure gene activity, useful when DNA is the only material available. Whereas RNA can only be harvested from fresh tissue under ideal circumstances, DNA is relatively stable and can be isolated from paraffin-embedded tissue. Although RNA is tightly regulated and variations in its abundance clearly relate to changes in cellular behavior, DNA is present in consistent concentration in all diploid cells. DNA-based genomic assays

exploit the pathologic exceptions to this rule, in which increased or decreased amounts of DNA are observable. When associated with disease, alterations in DNA copy number, in turn, likely correspond to changes in RNA through the phenomenon of gene "dose." Cancer provides the best examples, in which many measurable genetic abnormalities play a role in tumor genesis, including translocations, chromosome loss (partial or whole chromosome), and chromosome gain. The ability to survey across the genome for loss of DNA regions containing tumor suppressor genes or gain of DNA regions with oncogenes is of great interest to both clinicians and researchers.

An early DNA based genomic assays called comparative genomic hybridization (CGH) was first described in 1992 in the journal *Science (61).* The technique involves first labeling reference and test DNA samples with different fluorescent dyes (red and green) and then hybridizing to normal metaphase chromosomes. The relative fluorescence between test and reference sample at different loci along the chromosome is proportional to the relative number of DNA copies. In this way, amplifications of whole chromosomes or portions of the chromosome are reliably detected, although the resolution of these regions is low, in the range of 20 Mb. Once a locus of interest has been identified, investigators can attempt to identify genes within it by fluorescent *in situ* hybridization (FISH). As the resolution of the method is quite coarse, gene identification through CGH can be a laborious process.

In 1998, Pinkel et al. described a microarray-based technique that improved the resolution of CGH dramatically from 20 Mb to 40 kb *(62)*. Instead of using metaphase chromosomes as the probes, the authors used a library of genomic DNA, each sequence mapped to a known location on a chromosome. The genomic DNA probes are spotted on a glass slide, as were the cDNA probes from the microarray example in Section 3.2. Also similar is the process of labeling test and reference DNA with different fluorescent probes, hybridizing against the microarray, and scanning for relative fluorescence. In this manner, individual chromosomes and in fact the entire genome can be evaluated for areas of amplification or deletion with a much finer resolution than earlier. As with CGH, a DNA region of importance is highlighted by array CGH, not individual genes. At this resolution, however, using genomic sequences, an investigator can narrow candidate genes to a relatively small number for further investigation. Refinements continue to be made to this technique to increase the resolution of the assay *(63)*.

An alternate DNA microarray-based genomic approach is the SNP (pronounced "snip") array. The most frequent variation in the human genome is a single base substitution called a single-nucleotide polymorphism (SNP). To be considered a SNP, and not a mutation, an event frequency should be >1%, although many occur with much higher frequency. As expected, most SNPs occur in nontranscribed portions of the genome, where they would be thought to have little biologic significance. Although these nontranscribed SNPs have no functional significance, they provide a useful marker for genetic investigation in the following manner. A somatic cell is sampled to determine the germline genotype at a SNP locus. If the cell is heterozygous for the SNP, then the site is potentially informative in evaluating a tumor for DNA loss. In other words, if the somatic cell is heterozygous for a SNP and the tumor cell is homozygous, this suggests that one chromosome or a part of one chromosome containing the missing allele has been lost. Loss of genetic material in this way is termed "loss of heterozygosity" (LOH). Homozygous SNP loci in the somatic cell provide no information on LOH in the tumor cell.

Affymetrix has produced versions of its oligonucleotide array that uses these principles to identify LOH in tumor cells *(64,65)*. The current version contains probe sets to genotype approx 10,000 human SNPs selected for their high frequency of heterozygosity and even coverage of the genome. The chips are prepared in exactly the same manner as those described in Section 3.5, although the probe design strategy is different. Probes are again 25 bp in length, with the base in position 13 representing the SNP. One probe for each possible substitution at the SNP site is included on the chip as well as probes for flanking sequences of DNA. In this way normal somatic cells can be assayed to determine the loci at which an individual is heterozygous. Tumor cells are assayed and the results compared. Heterozygous loci in the somatic cell that register as homozygous in the tumor cell indicate areas of interest in tumor biology, especially when they occur in multiple samples. The pattern of LOH itself might be useful for identifying tumor subtypes. Areas of LOH can highlight the location of tumor related genes, such as tumor suppressor genes or oncogenes.

7.3. OTHER ARRAYS Having described two examples each of DNA and RNA genomics assays, we acknowledge that many other similar technologies are in various stages of development. We have focused only on those with the most immediate clinical applications and that are perhaps the most mature. Additionally, we have omitted a wide variety of microarray-based technologies that are not, strictly speaking, genomics techniques. Examples of these include the tissue microarray, in which small tissue samples from many sources are assembled and displayed on a single glass slide *(66)*. Tissue arrays provide an efficient method for measuring many samples for features such as protein or DNA mutations. The slide can be stained for the characteristic and scanned for its prevalence across the many samples on the array *(67)*. Other powerful DNA microarray technologies have been developed for the purposes of mutation detection, sequencing, genotyping, and polymorphism detection. These are, again, not generally genomics tools and we refer the reader to the original publications and reviews *(68–71)*.

8. PROTEOMICS

When the term "genome" was coined in the 1960s, it was applied specifically to collections of genes, the physical unit of heredity, and the DNA that comprised them. Because RNA's major function is to translate genes into proteins and because RNA is named for its corresponding DNA sequence, it is convenient to speak of RNA as a component of the genome and expression genomics. The same logic might have extended to the polypeptide sequences dictated by RNA. Recognizing that expressed protein measurement offered challenges beyond those previously encountered in gene expression, a new term

was coined in 1994, "the proteome" *(72)*. A brief discussion of the additional challenges posed in the study of protein systems helps contextualize the pace of functional proteomics relative to functional genomics.

In Section 1, we discussed the paramount importance of sequencing the human genome for genomics analysis. DNA sequencing has direct relevance for proteomic analysis, in that polypeptide can often be traced to originating DNA sequence. However, proteins, unlike DNA and RNA, undergo extensive posttranslational modifications. Modifications in many cases dictate the biologic properties of interest, independent of polypeptide sequence. It has been estimated that there, on average, 10 different forms of any given protein, each with different properties. It is a combination of the relative and absolute abundance of the protein forms that dictates biologic function. Although there are protein databases similar to those for genes, there is nothing akin to the Human Genome Project that would allow for automation in the way that has been done for DNA and RNA.

In addition to the existence of multiple posttranslational forms of a given protein, protein structure is more complex than that of DNA or RNA. Nucleic acids are composed of only 4 bases sequenced in linear fashion. Proteins are composed of almost two dozen amino acids in which secondary, tertiary, and quaternary structure often dictate function. In summary, whereas genomic analysis focuses essentially on the quantitative expression of genes, the field of proteomics has a broader set of challenges *(73)*. Current proteomics research can be divided into three distinct categories to address these challenges *(74–76)*. The first is abundance proteomics: a description of the relative and absolute abundance of different protein species in a given biologic system. The second is cellular proteomics: the study of protein–protein interactions and the function of protein networks. Finally, structural proteomics refer to the work of identifying novel proteins and their constituents. Although initial progress in all three areas is promising, no area of proteomics currently approaches the clinical usefulness of expression genomics, as we will see.

The most commonly used tool in the field of abundance proteomics is the technique termed "two-dimensional polyacrylamide gel electrophoresis (2D PAGE). First described in the 1970s, 2D PAGE separates proteins along two axes—first by pH gradient to sort by charge, and then by electrophoresis to sort by molecular weight. A heterogeneous mixture of proteins from eukaryotic cell lysates can reproducibly resolve up to 10,000 proteins, as seen in Fig. 8. The gel can be digitized and analyzed by software to compare similar protein spots across experimental conditions in a manner similar to other array-type experiments. This technique has been particularly promising in discriminating proteins that might serve as new tumor-specific markers that can be characterized by the approaches presented below.

In the area of cellular proteomics, investigators have employed a microarray experimental design similar to the cDNA array to describe novel protein–protein interactions *(77)*. On a protein microarray, polypeptides are produced in a high-throughput system such as a yeast vector. Individually identified proteins are spotted onto a glass medium by a robot in a manner similar to the production of a cDNA array described in Section 3.2. Once the chip is produced, a single labeled protein can be placed in solution, hybridized against the chip, and scanned just as with the DNA microarray. In this way, areas of fluorescence suggest protein interaction networks that can be explored by other methodologies.

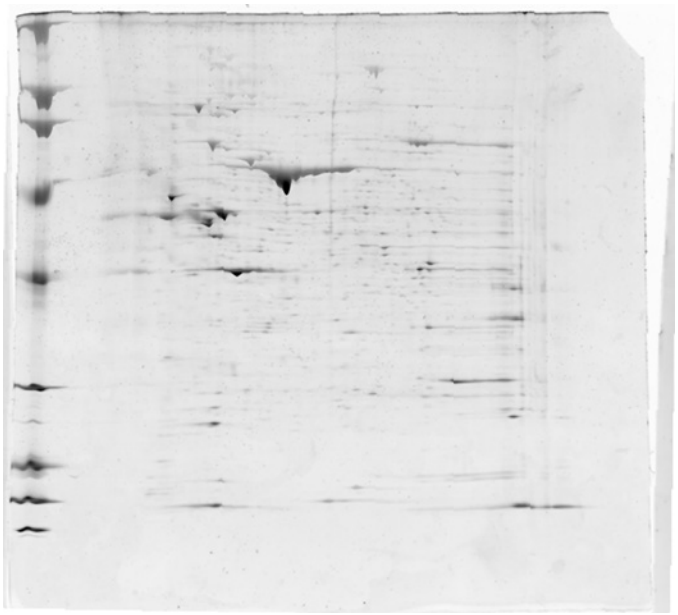


Fig. 8. 2D PAGE. (Courtesy of Liza Makowski, Ph.D., Division of Metabolic and Complex Diseases. Harvard School of Public Health, Boston, MA.)

Perhaps most exciting in the broader field of proteomics is its potential to identify proteins through structural proteomics. The traditional approach to protein identification has been to clone the gene encoding a protein of interest, a famously laborious task. In the genomics era, however, evolving techniques such as mass spectroscopy have linked physical properties of proteins directly to known DNA sequence information. In the most powerful examples, these tools can be automated for large-scale protein identification. In mass spectroscopy, a sample, such as a protein, is converted into a charged gas. The charged-gas particles are accelerated by an electric field into a linear detector. The time of flight in that linear detector can be measured, a value proportional to the mass of the particle *(78)*. To use mass spectroscopy to identify an individual protein, the protein is isolated by standard methods and digested by an enzyme such as trypsin to generate a reproducible set of peptides of varying lengths. This set of peptides will produce a highly reproducible pattern when analyzed by time-of-flight mass spectroscopy. If the protein has previously been described, the pattern can be matched. For a novel protein, a specific pattern can suggest a theoretic peptide sequence by matching particle mass to likely peptides with that mass. The theoretic peptide sequence can be related to a corresponding DNA sequence from the genome *(79)*.

This brief introduction to proteomics is mostly to suggest a range of potential uses for proteomics techniques in biomedical research and, ultimately, for patient care. In terms of clinical uses, however, we have not offered any specific examples. The

lack of clinical examples is not an accident, as proteomics remains exclusively a research tool at this time.

9. ANALYSIS OF DNA MICROARRAYS AND STUDY DESIGN

For the remainder of the chapter, we return to the DNA microarray platforms developed in Section 3 and, in particular, the oligonucleotide platform. Although there are significant differences between the cDNA and oligonucleotide technologies, the underlying assay principles are the same as highlighted in Table 1. The similarities extend to the analysis and interpretation phase of the assay, although, again, there are differences in implementation by platform. Given that this text is not intended to prepare the reader to perform analyses, but as an introduction for the interpretation of microarray experiments, we will limit our further discussion to the oligonucleotide format only.

Microarray assays are generally employed for only a limited number of study designs. Relatively infrequently, a time series design is used in which a single subject or group of subjects are sampled at a variety of time-points or experimental conditions (such as increasing drug concentration). Gene expression is followed over time and concentration either with respect to a specific outcome, such as differential response to stimulus, or to determine patterns of gene regulation. Such experiments have been useful in demonstrating the potential of microarrays for elucidating complex molecular networks, even when the function of all genes is unknown *(80)*. Time series comprise a minority of DNA microarray experiments, many of which are directed at molecular pathway elucidation rather than clinical questions, and we will not consider them further.

The most common experimental designs sample subjects only once, not serially, and address one or both of the following questions: (1) What are the dominant patterns of gene expression in this sample, without regard to any specific outcome or phenotype? (2) What is the dominant pattern of gene expression with regard to a specific outcome or phenotype? An investigator interested in discovering previously unrecognized tumor subtypes is essentially asking Question 1, for example. In a different analysis the investigator trying to define genes associated with aggressive cancer might ask Question 2. Finally, an increasingly common experimental design is the validation of results obtained in the process of answering Questions 1 and 2 *(12,81,82)*. Before addressing the methods needed to execute these analyses, we will consider the final data processing and filtering steps.

9.1. NORMALIZATION The most common DNA microarray experiments attempt to make meaningful comparisons of gene expression patterns across samples. A meaningful pattern is one that relates to underlying biology; yet, patterns could emerge that relate either to experimental conditions or error as well. Proper experimental design and analysis are the tools needed to minimize the influence of error and experimental conditions. We have discussed elements of experimental design in Section 5 in this regard and now turn to the role of experimental conditions. The process of accommodating for experimental conditions in order to elucidate meaningful patterns in the microarray data has been given the name normalization and we will introduce its important features here.

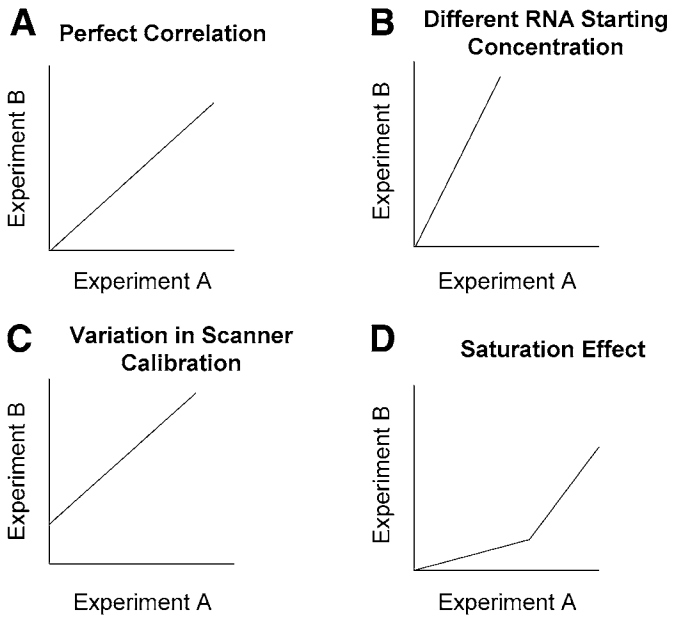


Fig. 9. Variation in expression values by experimental conditions.

Microarray images contain a variety of perturbations that are solely related to experimental conditions, and if not corrected for, they will result in uninformative patterns in the final gene expression analysis. Figure 9A–D shows examples of how gene expression values can be altered by differences in starting RNA concentration, variation in scanner calibration, and saturation effects, all without differences in sample biology. Each part shows results of paired experiments in which RNA from the same source has been divided and hybridized on two separate chips. The graphs show the range of expression value in experiments A (without regard to specific probes) vs the validation in experiments B assuming different experimental conditions.

Figure 9A shows a perfect correlation of expression values across the range of expression on the chip—the ideal situation in which experimental conditions were exactly the same across the two assays. Fig. 9B shows the results in which the RNA aliquot was divided into unequal parts 1/3 (experiment A) and 2/3 (experiment B). Expression at every point in experiment B will be twice that of experiment A solely because of starting RNA concentration. When the investigator is aware of unequal RNA starting concentrations, the perturbation is easily corrected by dividing all expression values in experiment B by the appropriate constant—2 in this case. In Fig. 9C, we see the influence of systematic variation caused by the scanner calibration, in which every probe in experiment B has expression systematically increased by a constant. This effect is corrected by subtraction of that constant from all probes in experiment B. Finally, we see the influence of a saturation effect in Fig. 9D, in which small changes in experiment A correspond to larger expression changes in experiment B above a threshold brightness. Although more difficult to recognize, saturation effects can also be corrected by relatively simple computational methods. Although presented as isolated phenomena in this example, in actual practice all

three effects are likely present in combination for any given two-array comparison.

The optimal normalization technique to address effects shown in Fig. 9 is an area of active research and most of the analytical software contains some form of normalization procedure *(83)*. In the simplest case, a procedure called standardization can correct the effect (Fig. 9B) by solving the equation $Y = MX$ for the slope M. It is similarly trivial to use linear regression to solves the equation $Y = MX + B$ for the slope and constant and normalize accordingly (useful to address combinations of effects from Figs. 9B,C). Figure 9D requires a more sophisticated modeling to address differential effects that occur across the range of expression. A family of regression techniques of the type commonly know as LOWESS regression or locally weighted polynomial regression have been employed for this purpose *(26)*. Another common approach is to apply probabilistic distributions to all probes in the experiment. For example, the mean expression value can be set to 0 and standard deviation to 1 for all probes on an array and all arrays in an experiment. Each of these methods has its strengths and weakness; for example, the probabilistic method assumes a Normal distribution of expression values, when, in fact, the distribution may not be known. In general, the relative strengths and weaknesses of normalization methods are beyond the scope of this review, however, we emphasize that some form of normalization should be a part of any current array analysis.

The normalization techniques described thus far have the underlying assumption that those probes with high (or low) expression in experiment A are the same in experiment B. Normalization in this case only serves to recalibrate experimental conditions, not alter the underlying biologic relations. This is a reasonable assumption when experiments A and B use RNA from the same source, however, most microarray experiments compare RNA from different sources. When RNA is isolated from different sources, we assume that they will have different RNA expression patterns that relate to differences in the biology of the sample. The most highly expressed RNA from sample A will be different from sample B, so that a normalization curve such as that seen in Fig. 9 is much more challenging to construct. When the probes that measure expressed RNA at the highest (lowest) levels are different across the normalization procedure, normalization reduces the very biologic variation that we are interested in studying. It might surprise the reader to realize that of all the challenges posed by microarray technologies, it is this problem of normalization that remains among the most difficult in bringing the technology to clinical use *(83)*.

Rank-invariant normalization is one of a number of techniques that has been suggested to better maintain biologically significant expression difference while performing adequate normalization *(84)*. In rank-invariant normalization, all genes on each array are ranked by the level of their expression. In theory, many genes unrelated to the biologic process of interest will be expressed on each of the two arrays at a wide range of expression levels. For example, some housekeeping genes are expressed at consistently high levels in all cells. These genes might have different absolute expression level according to the experimental conditions, yet their ranks should be similar across arrays. Rank-invariant normalization selects sets of genes with similar ranks across arrays, at high, medium, and low rank, for example, and performs local regression similar to the LOWESS method. In this way, only genes that are invariant across arrays provide information on the normalization scaling factors, yet all genes are normalized. Initial evidence suggests that this method is quite promising.

Other normalization techniques have been employed, both computational and experimentally based. A number of authors have tried spiking RNA species of known concentration in all samples across the experiment and the use of housekeeping genes to normalize in a manner conceptually similar to the rank-invariant technique. In practice, these methods have not proved successful. In summary, normalization is a vital step in making meaningful comparisons across microarray experiments. There are a wide variety of techniques, and although some are perhaps more promising, no consensus has been reached on the optimal method *(85)*.

9.2. HYPOTHESIS TESTING Normalized data are ready for hypothesis testing, however; the most familiar tool for this process often might not be appropriate for genomics applications. Classical biostatistics has evolved a methodology for addressing the challenges of hypothesis testing and study design, the basic principles of which are familiar to most readers. In brief, the method works as follows. An investigator wishes to demonstrate that a parameter A is greater than B, in which the measurements of A and B are associated with a certain error. The investigator collects enough samples of A and B so that the measurements of error are smaller than the expected difference between them. To prove that A > B, the investigator states a hypothesis that he wishes to reject (usually a null hypothesis), such as A is equal to B, and an acceptable threshold for making an error in the conclusion. By convention, the threshold is usually a 0.05 chance of stating that the null is false when, in fact, the null is true. In other words, there is a 5% chance that A and B are equal when we say they are not. If we were to perform the experiment twice, we would increase the chances of making that same error to 0.1 ($0.05 + 0.05 = 0.1$). In summary, using classical statistics to most efficiently prove that A > B, an investigator needs to collect many samples of A and B and test his hypothesis only once. This is the form in which most clinical trials are conducted, where a large cohort of patients is assembled to test one hypothesis, such as the effectiveness of a drug therapy.

Consider the case of most genomics experiments in light of the previous example that accrue small numbers of subjects, rarely more than a few dozen, while collecting data on thousands of genes. Each of the thousands of genes represents a potential hypothesis, a worst-case scenario by classical statistics methods. With few samples, individual genes cannot be measured reliably. With many hypotheses to test, the number of false positives will be large. For example, when 10,000 genes are analyzed at the 0.05 threshold, we would expect 500 false-positive results ($10{,}000 \times 0.05 = 500$) in addition to any true positives. Reducing the threshold to 0.01 might reduce the problem of false positives; however, this is at the cost of increasing false negatives. Although genomics data could be analyzed using classical biostatistics, in which several dozen samples were

analyzed for a single gene on the array, this represents a highly inefficient study design. The previous examples are grossly simplified, as there are adaptations of common statistical methods that overcome some of the shortcomings described, although there is general consensus that efficient analysis of microarrays requires novel analytical techniques *(86)*. Many of these methods are still under development, and unlike classical biostatistics, the conventions are not yet well established. Despite the ongoing evolution, an analytical framework with broad applicability has emerged that will be discussed in the following subsections.

9.3. GENE FILTERING Thus far, we have focused on processing microarrays without regards to the specific genes on the arrays. A shorter list of useful genes statistically and functionally associated with the outcomes of interest might be preferable to the indiscriminant use of all genes on the array. Accordingly, most investigators try to eliminate genes that have no plausible association with the disease states of interest or are likely to contribute error to the analysis. Attempts to do this systematically by manual review of genes, especially across multiple samples, are hindered by the shear volume of data. As an initial step to decrease the chances of spurious results by data reduction, most analyses include a component of gene filtering. Gene filtering is the process of removing from further analytical consideration a large number of genes unlikely to contribute to the results or have undesirable properties for the specific hypothesis at hand *(87,88)*. As suggested in the previous subsection, no consensus exists for the optimal filtering of genes, but we can consider the most commonly used approaches. Although gene filtering might be useful, as with any step in the preprocessing of the data, it likely influences the results of the final analysis. In this regard, investigators and readers should take note of the filtering method and stringency of the criteria used.

The usefulness of genes is often considered in the context of the entire sample set. Genes that do not vary across samples are generally not considered useful and can be removed from further analysis. Lack of variability can relate to the biology of the sample set, the quality of the probes, or other factors. When there is no variation, however, genes can only be thought of as contributing noise without signal. The variability of a gene can be described in a number of ways, such as the percentage of samples in which the gene exceeds a threshold expression value or the ratio of standard deviation to mean across samples. Eliminating the large number of genes with expression near zero often results in the greatest data reduction. Although gene variability across samples is a desirable analytic characteristic, a gene marked present in one or few samples out of many raises red flags. Although a gene expressed in only one sample can represent a biologically important finding, it often represents a false positive,which prudence suggests striking from further analysis.

A number of authors have proposed using replicate microarray, in which the same sample is hybridized against two different arrays to detect the probes with the highest reliability, filtering out low-reliability probes *(14)*. In addition to filtering based on gene distribution, strategies can be based on other gene properties such as underlying biology. Investigators can generate gene lists that relate to biologic processes such as developmental or inflammatory and systematically include or exclude them from analysis.

9.4. GENE SPACE In place of probabilistic distributions, the concept fueling classical statistics and hypothesis testing, the current paradigm proving most useful in the analysis of microarrays is that of multidimensional space or alternatively gene space. An example shown in Fig. 10 serves to highlight the basic principle. Suppose that that we perform 10 microarray experiments on 10 separate samples. Each microarray originally surveys 10,000 genes, but after processing and strict filtering, we are left with only two genes for the analysis. These two genes represent a two-dimensional space shown here in Fig. 9. Each microarray experiment, in turn, can be placed in that same two-dimensional space according to the expression levels of genes 1 and 2. In this manner and for this analysis, each microarray is represented fully by one point in this theoretical two-dimensional space. Extending the model to three-dimensional space, a third gene would define a discrete location in three dimensions. Although conceptually abstract, it is computationally trivial to extend the model indefinitely into the *n*th-dimensional gene space, in which the expression levels of *n* genes defines a unique point in gene space.

Consider some advantages of the gene space model. Many samples, each with thousands of data points (gene expression values), can be represented by a single point in space. In place of describing the interrelations of thousands of genes over multiple samples, the analysis can be framed in light of the distance of one array from another in gene space. Figure 10 illustrates this point, where the vectors **(A)** and **(B)** represent distances between samples and between groups of samples, respectively. Accordingly, vector **(A)** suggests a relatively short distance among samples in a group compared to a larger vector **(B)** between the two groups. Any two-sample comparison is reduced to a single distance measure no matter how many genes or dimensions of gene space are involved. Using the concept of distance, the problem of multiple hypothesis testing described in Section 9.2 is averted and we have the most useful current model for the analysis of microarray experiments.

Although a useful concept, distance measures have limitations in modeling biologic systems. First, any vector such as vector **(A)** from Fig. 10 has value equal to $[(\text{Gene1}_{\text{array1}} - \text{Gene1}_{\text{array2}})^2 + (\text{Gene2}_{\text{array1}} - \text{Gene2}_{\text{array2}})^2]^{1/2}$, as described by the Pythagorean Theorem. The straight line measuring distance between two points is a measure known as Euclidean distance and is only one of many ways in which distances can be described. For example, we might choose to weight-relative gene contributions differently or use other well-described distance measures *(89)*. Additionally, there are cases in which distance might be ill-defined by any measure, such as a dichotomous variable like gender or "on/off." Ultimately, the relevance of distance measures and the models that use them is reflected by their performance in biologic systems. As we will see, they often perform quite well, and when they do not, there are other model systems.

9.5. MACHINE LEARNING In Section 5, we noted that the two most common microarray experiments attempt to answer the following questions: (1) What is the dominant pattern of gene expression in this sample, without regard to any

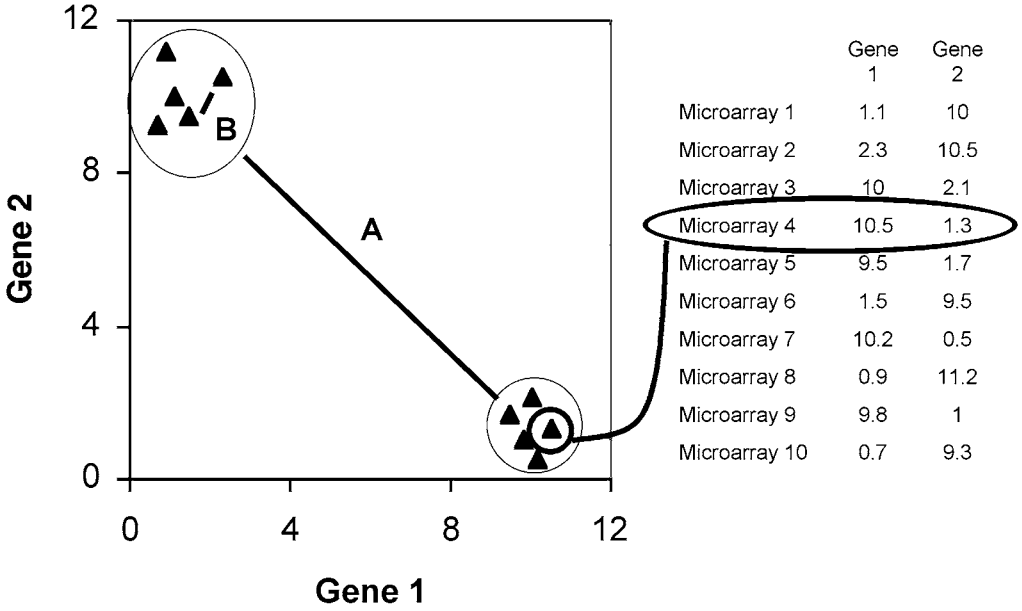


Fig. 10. Two-dimensional gene space.

specific outcome or phenotype? (2) What is the dominant pattern of gene expression with regards to a specific outcome or phenotype? The search for patterns defined by multiple data points has been an uncommon problem in biomedical research, which traditionally addresses hypotheses singly by hypothesis testing. Researchers in other field of science such as artificial intelligence, however, have been more interested in problems of pattern recognition *(90)*. The following subsections will develop some basic tools of artificial intelligence that have been adapted for gene expression arrays. Our focus will be more on method application without rigorously detailing the computational aspect.

The technique of finding patterns in data without regard to specific outcome, (Question 1) has been called unsupervised learning and clustering. For example, given a group of tumors that appear similar by current diagnostic methods, are there any gene patterns that represent unrecognized biological differences? In this example, we are looking for gene patterns without regard to any outcome, such as tumor behavior or patient survival. Using the language of machine learning, neither the classes into which we wish to divide the samples nor the genes needed to make those divisions are known at the start of the analysis. Alternatively, techniques for addressing Question 2 have been called supervised learning or classification. Continuing the undifferentiated tumor example, when we look for genes in these samples that associate with a particular outcome, such as aggressive vs indolent behavior, the methods used are those of supervised learning. In supervised learning, the classes into which we wish to divide the data are known, but the genes used to accomplish the task are not.

9.6. UNSUPERVISED LEARNING

9.6.1. Introduction Commonly used unsupervised learning techniques fall broadly into two categories (hierarchical and partitioning), of which hierarchical clustering is more widely used *(90)*. The preference for hierarchical clustering is at least twofold after considering the following challenges of machine learning. Computational resources to solve many machine learning problems are not currently available. The chasm is broached by abandoning the search for an exact solution, accepting, instead, an approximate solution. One approximation is to run a complex procedure iteratively until it converges near a single value. Alternately, instead of evaluating all possibilities, the algorithm selects a random solution, which is iteratively optimized. In these approaches, running additional iterations or choosing a different random seed might result in a different solution, although, in general, the approximations are close. Hierarchical clustering, or at least one version of it, avoids the uncertainty by providing a method that can be solved explicitly that does not rely on a random seed. The general preference for hierarchical clustering provides additional practical incentives, for its use in that software is widely available.

9.6.2. Clustering and Dendrograms Distance measures suggested in Section 9.4 are the basis for hierarchical clustering. Recalling that a sample is uniquely identified in multidimensional gene space by a single point, it is computationally trivial to calculate a distance between any two samples. Samples for which the between distance is smaller have expression profiles that are more similar than those with larger distances. In this way, some samples might be in close proximity, clustered in other words, compared to other samples. The two broad types of hierarchical clustering are termed "agglomerative" and "divisive" *(91)*. Divisive clustering is a top-down approach in which all samples start in a single group and are serially divided into smaller groups until each sample is its own group. Computationally, divisive clustering is difficult to solve explicitly for large numbers of samples. For this reason, agglomerative clustering is preferred. Agglomerative clustering is a bottom-up approach in which individual samples are serially grouped to build progressively larger clusters. In this method, each sample is paired with its nearest neighbor to build

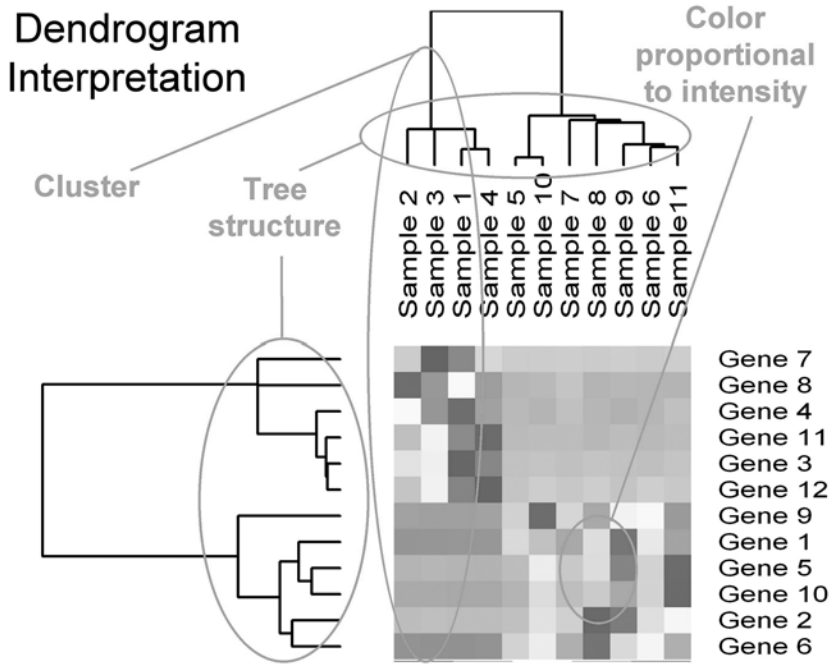


Fig. 11. Dendrogram. (Figure appears in color in insert following p. 172.)

the first layer of clusters. The newly formed cluster is assigned its own point in gene space that is in the neighborhood of the two samples which comprise it. For example, a line in gene space can be drawn between the two samples and the cluster location assigned to the line's midpoint. Once new locations have been calculated for all pairwise clusters, a second round of distance measures between clusters can be calculated. Clusters that are closest are merged and the process repeats until all samples have been agglomerated into one group.

Hierarchical clustering results are displayed in the form of a dendrogram. Although there is some variation in dendrograms format, Fig. 11 displays many of the typical features. Columns in this example represent samples and rows represent genes. At the intersection of a row and column is a colored rectangle whose intensity is proportional to the gene expression in that sample. Depending on the software, a threshold gene expression level, such as average expression, is often calculated in order to further use color for display purposes. Samples in which expression is above and below the threshold are displayed in different colors, with color intensity proportional to the distance from that threshold. In this example, red represents greater than average expression and blue represents less than average. The dendrogram represents the clustering of both samples and genes in separate tree structures seen above and to the left of the grid, respectively. Samples (or genes) closely related in gene space are placed adjacent to each other in the tree and joined with shorter branches. In Fig. 11, samples 1–4 and genes 3, 4, 11, and 12 are very closely related, forming clusters with short branches. Less closely related samples (or genes) have longer branches and are located farther apart in the tree.

Dendrograms are a convenient method for the display of microarray data from multiple samples and are widely used tool for viewing experimental analysis. Although useful, they should be viewed with the following shortcomings in mind. Once a sample has been assigned to a branch, it cannot be reassigned, meaning an error propagates through subsequent branching decisions. Branch lengths, although proportional to distances between samples, are an ambiguous measurement on which to judge the strength of a cluster assignment. For these reasons, it can be difficult to judge where cluster boundaries occur or if any significant clusters are present. Fig. 11 suggests two strong clusters, samples 1–4 and samples 5–11, as defined by the major branching of the tree. It is less clear whether these strong clusters should be further subdivided and where those divisions might be.

9.6.3. Partitioning Methods *K*-means clustering is one of the most commonly described partitioning method (as described in Fig. 12) and again relies on gene space and distance measures *(92)*. The user first selects a number of classes into which the samples should be divided, two in the example. The algorithm selects a point in gene space, called a centroid, for each of the classes. The choice of starting point for the centroid might be random or based on some feature of the data depending on the exact method. In the first iteration of partitioning, each of the samples is then assigned to the class of the centroid to which it is closest, as shown in Fig. 12A. Using all samples assigned to a given centroid, the geographic center of their positions is calculated, and the centroid is moved to that location. Samples are reassigned to a centroid based on its new location, repeating the iterative process until either the centroid's position remains stable or other analytic parameters have been satisfied.

Partitioning methods suffer from the need to use iterative approaches and random seeds (random starting locations for the centroid). Placing the centroids at a different starting location or running a different number of iterations could result in variations in the final cluster assignments. In addition, most procedures require that the user specify the number of classes

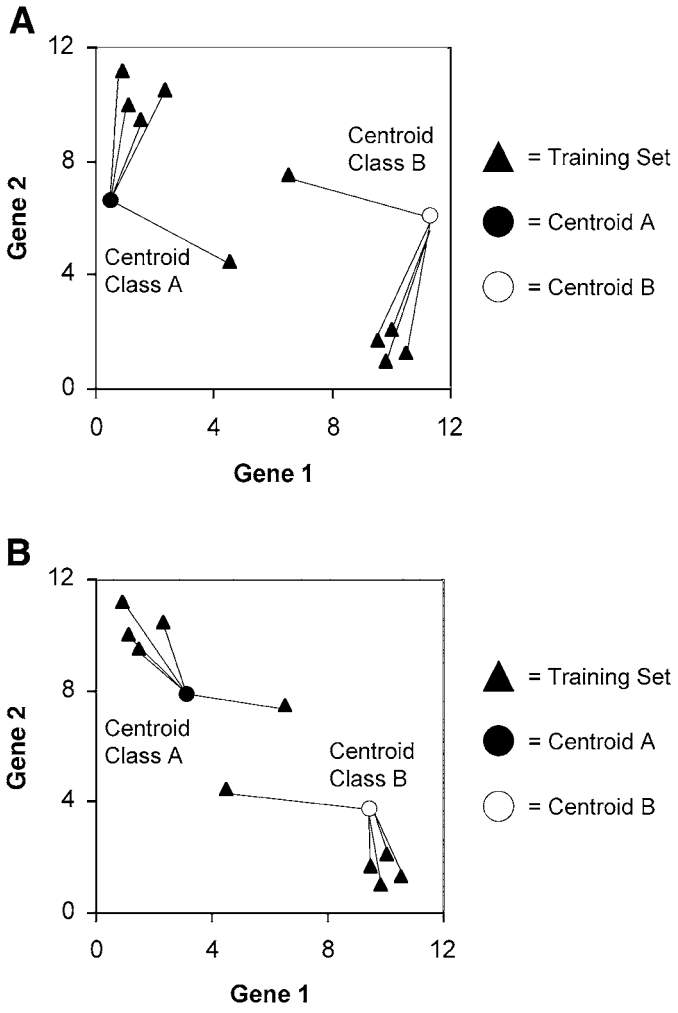


Fig. 12. *K*-Means clustering: **(A)** iteration 1; **(B)** iteration 2.

into which the sample should be divided. Because the classes themselves are unknown in the setting of unsupervised learning, it is reasonable to expect that their number would also be unknown. For example, in attempting to describe new tumor subtypes, it might be unclear how many subtypes to expect.

9.6.4. Which Method to Choose? There is great interest in improving the performance of unsupervised learning algorithms and refinement of specific aspects, such as error description. Work continues on existing hierarchical and partitioning method, as well as the development of novel techniques. In the absence of consensus, an extended discussion of the relative merits of individual applications is unlikely to be fruitful for this audience. Nonetheless, researchers and clinicians alike require an interim objective measure to asses the quality of the analysis. We suggest the following framework in the context of ongoing development. The study should clearly delineate the hypothesis being tested in terms of supervised learning, unsupervised learning, or other analysis. One quality of an unsupervised learning experiment is that, by definition, it is hypothesis generating and the classes described by the analysis are unknown *a priori*. Because the classes are unknown, the criterion which to judge the quality of the algorithm should not be its ability to determine known classes. Better criteria include such features as reproducibility of clusters or measures of cluster strength. The hypotheses generated in an unsupervised analysis can be verified by supervised analyses or other method.

Within the context of a stated hypothesis, any specific study goals that might clarify the choice of one analytic approach should be noted. For example, if there is a strong *a priori* reason to only look at two subgroups in an unsupervised analysis, an author would have a reasonable argument to perform a portioning analysis with $k = 2$. If there are no such specific requirements, then the choice of one method over another should be made explicit or consideration should be given to the use of multiple methods. In the absence of other justification, hierarchical clustering should probably be viewed as the standard for unsupervised learning.

9.7. SUPERVISED LEARNING Supervised learning is the process of segregating samples into known classes based on collections of data called features. As in the unsupervised learning example, the features are genes. The machine learning techniques along with user inputs will decide which genes should be used and how. Classes can represent a wide variety of biology, from phenotype to clinical behavior. Examples of supervised learning include using genomic data to differentiate known tumor types, aggressive tumor behavior, response to drug therapy, or tendency to metastasize. Supervised learning algorithms have two components. The first component is a training stage in which a training dataset is used to determine the relationship between genes and outcome. The second component is a validation stage, in which the hypothesized relationship is evaluated in independent samples, also called the test dataset, not used to develop the model. In the validation step for which the classes are known, the model is evaluated for its error rate in predicting known classes using the selected genes.

In the parlance of machine learning, the process of finding a relationship among features, genes, and class assignments (e.g., clinical outcomes) is termed "building a classifier." The classifier itself is any set of rules to establish the relationship. Classifiers, their rules, and their subtypes take many forms *(93–98)*. In the simplest machine learning cases, a classifier predicts the class by relating responses to a series of binary yes/no features. For example, a training set of tumor specimens using a decision-tree classifier could establish a rule in which any female patient with an estrogen receptor-positive tumor would be assigned the to the class "breast cancer." In more complicated examples, every gene on a microarray might contribute a vote toward tumor classification, such that the importance of any single gene might be difficult to interpret.

9.7.1. k Nearest Neighbors The partial list shown in Table 2 shows that there are far more supervised learning techniques than unsupervised learning techniques. We will discuss two examples that have been widely used and then consider the wider choice of optimal technique. The *k*-nearest-neighbors technique is simple but useful, relying on the distance concepts previously discussed. A set of useful genes is selected from those provided by a training set of samples. The definition of what constitutes a useful gene is not explicit but, in general, implies genes that are differentially expressed in the different classes. Once these genes are selected, they

Table 2
Selected Supervised and Unsupervised Learning Algorithms

Class	*Subtypes*
Supervised	
Linear discriminant analysis	Weighted voting
Classification tree	Classification and regression trees
Density based	Naïve Bayesian
Regression	Predictive modeling
	Linear
Neural networks	Linear perceptron
Nearest neighbors	*k* Nearest neighbors
Support vector machine	Linear and nonlinear
Other	Bagging
Boosting	
Learning vector quantization	
Unsupervised	Random forests
Hierarchical	Agglomerative clustering (AGNES—agglomerative nesting)
	Divisive clustering
Partitioning	*k*-Means
	Self-organizing maps
Density based	
Model based	

define the multidimensional gene space that will be used for the analysis and each sample is mapped into their corresponding location. For our purposes, we select two genes defining a two-dimensional gene space that separates well three known tumor classes A–C, as seen in Fig. 13. Samples from the training set are represented by diamonds, with samples from similar classes segregating closely in gene space. To the extent that the expression of genes defining gene space describe the identities of classes A, B, and C, a new sample of those classes should also map in the same vicinity. The *k*-nearest-neighbor algorithm assigns the class of a new sample by selecting its *k* nearest neighbors, *k* representing a predetermined experimental parameter often 3 or higher. Setting $k = 3$ in the example, the hollow star would be assigned to class C because all three nearest neighbors are of class C. The solid star is of true class A, but it has two of three nearest neighbors of class B, suggesting that it would incorrectly be assigned to class B. In some formulations of the *k*-nearest-neighbors algorithm, the voting is not simple majority rule. For example, the nearest neighbor can be given the most weight or weighting can be proportional to overall distance.

9.7.2. Support Vector Machine An alternate classification approach used by the support vector machine classifier is to divide gene space into sectors which represent the classes. Again, the training set is used to select a set of genes that are useful for making class distinctions and samples are mapped into gene space. Two dimensions are shown in Fig. 14 for simplicity's sake, although, in practice, two dimensions are rarely sufficient to separate complex classification problems. Higher-dimension models often have 50–100 genes or more and are more likely to find gene spaces that separate classes well. Boarders around each class can be defined in two dimensions by a line, three dimensions by a plane, and higher dimensions by a hyperplane. Computationally, boarders are defined by classifiers such as the support vector machine algorithm, which maximizes the distance between samples and minimizes the distance that outliers fall into space defined by another class. Once defined, test samples can be mapped into gene space and assigned a class according to the sector in which they fall.

9.7.3. Which Method to Choose? Faced with the previous classification examples in conjunction with those enumerated in Table 2, the question arises as it did with unsupervised learning: Which method is preferred? In part, the answer is suggested by the work of Duda et al. in the form of the oddly titled "No Free Lunch" theorem *(89)*. In essence, the theorem states that there is no single best classifier, but rather the performance of a classifier is a function of the question being asked and the data available to answer it. For example, if method A is optimal at classifying a tumor into histologic subtypes based on a given set of genes, there is no guarantee that method A will work to classify those same tumors by clinical behavior. Similarly, given a different set of gene on which to classify, there is no guarantee that method A will still outperform other classifiers. The performance of a classifier in a specific situation can only be learned through validation. That is not to say that different types of classifiers might not be preferred in a given situation. For example, decision trees require function through a series of binary splits. Genomic expression data are continuous and might not be conveniently framed as a decision-tree problem. There are a number of considerations of this sort that enter into decisions when selecting an analytic approach and for which the reader should consider further study in classification *(99)*. For the reader of this text however, no matter what algorithm is selected, the performance should be evaluated in validation.

9.7.4. Validation The validation process takes a variety of forms. Genes that have been suggested as interesting are often validated by use of conventional techniques. The author can return to the original clinical sample and verify increased RNA expression by Northern blot, real-time polymerase chain reaction (PCR), or other method. In addition, investigators look for genes and proteins related to those suggested by the microarray analysis in order to prove overall consistency of the hypothesis-generating experiments.

In addition to validation of the biologic findings, the authors will usually present at least one form of analytical validation related to the classification component of the study. The method of validation can take a number of forms depending on the study limitations. The most common study limitation currently in the filed of genomics is that of sample size. An investigator's ability to validate results is significantly constrained by small sample size; however, a number of techniques have evolved for use even in the setting of very small studies.

Recalling that most microarray experiments contain relatively few samples, it is understandable and, in fact, efficient to use all of them in the process of building the classifiers. This produces a dilemma when trying to validate the performance of the model, as the classifier will always overestimate its true performance in the data that were used to develop it compared to independent data. The reason for increased performance in the training set is partially the result of to random chance in the

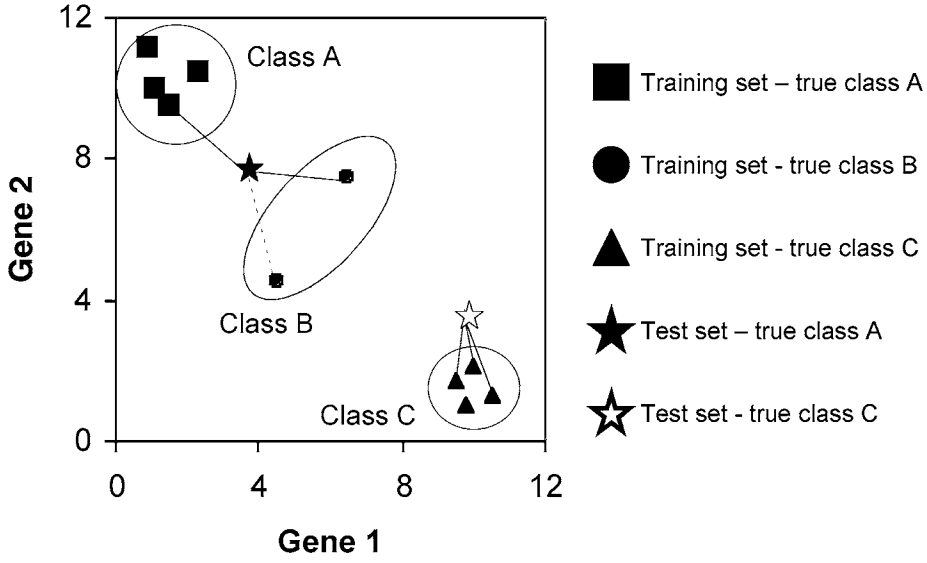


Fig. 13. *k* nearest neighbors.

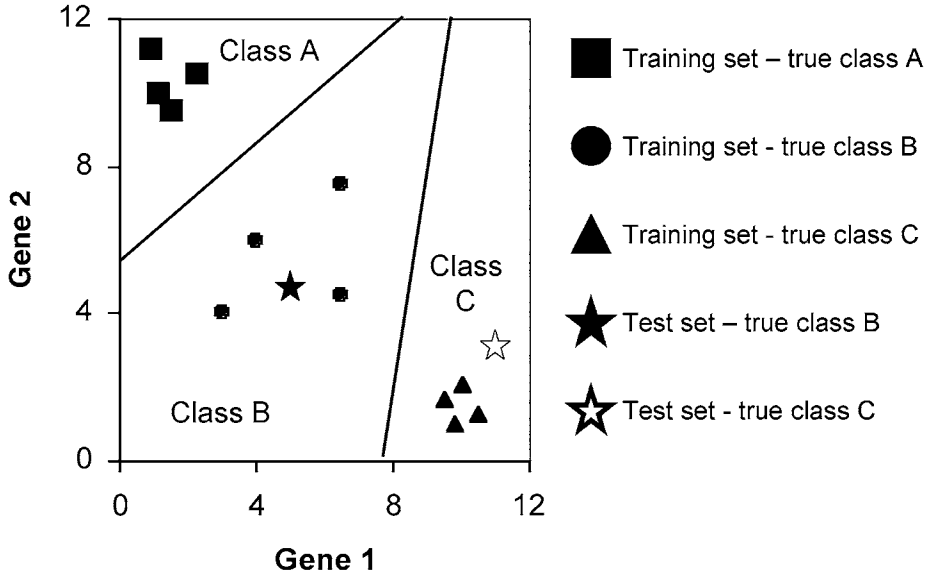


Fig. 14. Support vector machines.

measurement of gene expression. Just as the measurement of any given gene will vary somewhat around its true value because of measurement error and random chance, the measured group mean of a gene will vary around the true mean. To the extent that random error artificially increases the strength of a gene associated with a given class assignment, a classifier will select that gene preferentially. To the extent that random error artificially decreases the strength of association with class assignment, a classifier will omit that gene from the classifier. Because of random error, the classifier includes genes that looked better than they actually are for predicting a class, and it omits some genes that might have been included. To evaluate the extent of this and other shortcomings, classifiers need to be validated.

Of the techniques developed to overcome the limitations of small sample size, the most popular of which is probably the bootstrap. Bootstrapping, or sampling with replacement, is a method in which one or more samples is removed from the total sample set and, in their place, duplicate samples are substituted so that the overall sample size remains the same. For example, in a dataset containing samples A, B, and C, the bootstrap dataset would contain samples A, B, and B. Sample C is withheld for validation purposes.

Using the bootstrap sample set, the classifier is built and then validated on the samples that have been withheld. In this way, an error rate for the classifier can be calculated. The classifier that is made in this way is not exactly the same as the one that would have been made if the dataset were not modified, but it is very similar, and the error rate would be similar as well. The process is repeated multiple times, as many bootstrap datasets are constructed and multiple error rates are calculated. When the error rates are combined, we

have an estimate of the error rate for the final classifier, which is itself built on the total sample dataset, not a bootstrap. Note that the bootstrap validation is an estimate of the error rate for classifiers of a given type in a given dataset; it is not the true error rate of the final classifier in an independent dataset. An alternative to the bootstrap would be to withhold a proportion of samples from the training set to use as the test set, although this is generally regarded as an inefficient use of data. In another common alternative approach, a classifier that has been built previously can be tested on a prospectively collected data source.

9.8. LIMITATIONS AND SPECIAL CONSIDERATIONS OF SUPERVISED AND UNSUPERVISED LEARNING

We have already introduced the concern that distance measured in gene space has weaknesses as a foundation for both clustering and classification algorithms. Looking at the specific methods we present, we should again consider the implications of using distance. In all of the examples, we graphed genes 1 and 2 on the same linear 1–12 scale. In practice, however, genes can vary over vastly different gene expressions, and the distance algorithms will overweight those genes with higher expression. A gene that varies by a factor of 10, from 100 to 1000 for example, will be weighted the same as a gene that varies from 10,000 to 11,000 if a correction is not included in the analysis. Distance measures that correct for scale, error, or other considerations are available for use in all of the previously cited examples. In most cases, however, the selection of optimal distance measure will not be obvious to the investigator *a priori* and, again, might best be determined in validation experiments.

Although the concept of distance remains at the center of much of supervised and unsupervised learning, there are a many techniques listed in Table 2 that do not directly rely on multidimensional gene space. For the reader interested in further study, there are a variety of information sources to expound on specific methods, including the computational applications for those wishing to perform genomics assays themselves *(100,101)*. For those interested in evaluating the current literature however, we see little interest in full development of classification techniques at the current time. This is particularly true as the coming years we will likely see developments in the technology and, eventually, standards that are not readily apparent at this time. Accepting a lack of standard methods obliges us in some cases to view the machine learning component of some studies as a black box for the present time.

Ultimately, the black box of machine learning is no different that that of commonly used research methodologies and should not interfere with the ability to evaluate the overall quality of genomics research. For example, relatively few clinicians have training in linear regression, logistic regression, or Cox proportional hazards modeling, yet most have developed comfort with the results of studies reliant on the methods. By viewing genomics studies in the same modular fashion as those using regression techniques, we believe that the reader has sufficient information to evaluate the overall quality of the research regardless of the supervised learning methods used. The modules of a study or clinical application are exactly the section headings of this chapter, including study design, sample preparation, normalization, gene filtering, supervised/unsupervised learning, and validation.

10. STANDARDS

Interpretation of microarray experiments is not only the task of the investigator performing the experiment but also the scientific journal peer reviewers, the readers of those journals, and other interested parties. The scale of the genomic data in addition to its complexity has challenged all involved to seek new methods of communication in order to take full advantage of its promise. An early realizations was that print media is insufficient to allow full interpretation of these data-rich experiments. Most journals require that the primary data as well as associated experimental conditions and sample information be made available electronically via the Web as part of the publication process. The provision of data alone has often proven insufficient to allow meaningful scientific evaluation, including reproducing the authors results. Without standards for reporting results, elements necessary to interpret the arrays might not be included in the online databases *(10)*. For example, an author might publish only the raw image files without the final expression values or, conversely, publish expression without the image files.

Recognizing that reporting standards were needed the professional society, the Microarray and Gene Expression Data (MGED) Group initiated such a project *(102)*. Their initial work has produced a set of standards called the minimum information about a microarray experiment (MIAME) checklist (Table 3). The MIAME standards have been formulated to include the minimum information to interpret an array experiment and, as such, represent the information authors must provide upon publication in the scientific literature. There is evidence that these criteria are currently being implemented, although it remains to be seen if they will be widely adopted. The original MIAME criteria request a large amount of data per experiment without explicitly stating in what form it should be presented. In answer to formatting concerns, the MGED group has followed up with additional recommendations called the Microarray and Gene Expression Markup Language (MAGE-ML) and Object Model (MAGE-OM).

It is the microarray alone that requires standardization. The very feature measured by an array, the gene, is itself not a static concept. We discussed in Section 3.1 the need to link probe DNA sequence information to gene name, a process requiring standardized genomic databases. As gene sequencing began in earnest in the mid-1980s, it became clear that such central clearinghouses of information were necessary. One such database was created when the United States Congress funded the National Center for Biotechnology Information (NCBI) as a division of the National Library of Medicine at the National Institutes of Health in the mid-1980s. A central task was to catalog and publish genetic information to allow researchers to speak a common language, referring to specific DNA sequences in a standardized way. Consider the challenge previously facing scientists when discussing alternate splices of the same gene, where

Table 3
Summarized MIAME Checklist

Experiment design
- Type of experiment (i.e., normal vs diseased tissue, time course, others).
- Experimental factors: the parameters or conditions tested, (i.e., time, dose, or genetic variation).
- Number of hybridizations performed in the experiment.
- Reference used for the hybridizations.
- Hybridization design: description of the comparisons made in each hybridization, whether to a standard reference sample or between experimental samples; an accompanying diagram or table might be useful.
- Quality control steps (i.e., replicates or dye swaps).
- URL of any supplemental websites or database accession numbers

Samples used, extract preparation and labeling
- Origin of the biological sample and its characteristics (i.e., gender, age, developmental stage, strain, or disease state).
- Manipulation of biological samples and protocols used (e.g., growth conditions, treatments, separation techniques).
- Protocol for preparing the hybridization extract (e.g., the RNA or DNA extraction and purification protocol).
- Labeling protocol(s).
- External controls (spikes).

Hybridization procedures and parameters
- The protocol and conditions used during hybridization, blocking and washing.

Measurement data and specifications
- The quantitations based on the images.
- The set of quantitations from several arrays upon which the authors base their conclusions. Access to images of raw data is not required although preferred, authors should make every effort to provide the following: scanning hardware and software used, image analysis software, measurements produced by the image analysis software and a description of which measurements were used in the analysis, image analysis before data selection and transformation (spot quantitation matrices), final gene expression data table(s).

Array Design
- General design, including the platform type (spotted glass array, an *in situ* synthesized array), surface and coating specifications, and the availability of the array (the name or make of commercially available arrays).
- For each feature (spot) on the array, its location on the array and the ID of its respective reporter (molecule present on each spot).
- For each reporter, its type (e.g., cDNA or oligonucleotide) should be given with unambiguous characterization such as database refer ence and sequence
- For commercial arrays: manufacturer reference including catalogue number and references to the manufacturer's website
- For noncommercial arrays, the following details should be pro vided: source of the reporter molecules: for example, the cDNA or oligo collection used, with references, method of reporter preparation, spotting protocols, additional treatment performed prior to hybridization.

gene name does not describe a unique DNA sequence, but two or more sequences. The converse also occurred where multiple gene names described very similar DNA sequences when a gene was sequence and named by more that one researcher. To clarify these types of confusion, the NCBI initiated the GenBank database to which investigators submit sequence and other information and, in turn, have that information associated with a unique GenBank accession number. GenBank does not further evaluate the quality or veracity of genetic data, but rather serves as a repository. As GenBank sequence information grew, it became possible to perform automated analyses of the sequences. By evaluating GenBank sequences for overlap, it was possible to identify clusters of many entries that likely describe a single gene. In this way, a DNA sequence derived from several GenBank entries could be ascribed to a single UniGene cluster ID, again improving the ability of researchers to speak a common language.

With a common vocabulary in place for relating DNA sequences to naming and numbering systems, it becomes far simpler to discuss the function of individual genes and networks of genes. One of the benefits of standardizing databases such as GenBank is that automated analyses can be performed on vast quantities of data. In this manner, DNA sequences can be surveyed across species and potentially useful information from one species related to another. In the case of cellular metabolism in particular, it has been frequently noted that many genes are highly conserved across eukaryotic species. This realization has been used to link information from well-studied systems such as yeast to those where a given gene might not have been studied at all. In addition to automated information transfer such as this, there was a need to standardize species-specific information as it accrues, particularly with an eye for automating analyses of genomic experiments. For example, a microarray analysis might describe hundreds of genes that differ between samples—a quantity of information that is difficult to assimilate. The problem might be reduced in complexity if genes could be grouped by function, where a few groups of genes not hundreds of single genes, might be considered for their biologic significance.

Standardization of genetic information at functional levels is the interest of groups such as the Gene Ontology Consortium (GOC) *(8)*. The goal of the GOC is to create a dynamic controlled vocabulary that can be applied to all organisms even as knowledge of gene and protein roles in cells accumulates and changes. To this end, they have systematically annotated thousands of named genes from multiple species under three broad headings: biologic process, molecular function, and cellular component. Genes are assigned to a GO terms that describes its function in a hierarchical manner. In this way, a given gene can be described in terms of a very specific function as well as increasingly broader functions. For example, a gene might be specifically described as being involved in DNA strand elongation while also being recognized as a part of the broader headings of DNA-dependent DNA replication, DNA replication, and, ultimately, DNA metabolism. Additionally, a given gene can be annotated under all three headings; for example, a receptor can be located in the cell membrane (cellular component), participate in cell growth (biologic process), and have signal transduction properties (molecular function). In the case of a specific metabolic pathway, the gene ontology terms are a poor substitution for expert knowledge.

We have only mentioned a few agencies and databases that have contributed meaningfully to the analysis of genomic data.

Table 4
Selected Web Resources

Source	*Description*	*Web address*
National Center for Biotechnology Information	One of the single best sites; contains many datasets and array resources	http://www.ncbi.nlm.nih.gov/
Affymetrix	Wide variety of resources for their commercial products	http://www.affymetrix.com
Michael Eisen Lab	Software and other resources	http://rana.lbl.gov/index.htm
MGED	MIAME checklist, MAGE-ML and MAGE-OM tools	http://www.mged.org/
Dragon Database	Various bioinformatics tools for managing	http://pevsnerlab.kennedykrieger.org/dragon.htm
Bioconductor	Rich source of R programming language information; wide variety of tutorials and resources	http://www.bioconductor.org/
Kyoto Encyclopedia of Genes and Genomes	Variety of gene pathway information and other bioinformatics tools	http://www.genome.ad.jp/kegg/kegg2.html
Pat Brown's Lab and Stanford Genomics Center	Software and other resources	http://cmgm.stanford.edu/pbrown/ http://genome-www.stanford.edu/
Whitehead Institute	Software and other resources	http://www-genome.wi.mit.edu/cancer/
The Institute for Genomic Research	Software and a variety of bioinformatics resources	http://www.tigr.org/
ExPASy Molecular Biology Server	Home to multiple genetic and protein databases and bioinformatics resources	http://us.expasy.org/

The World Wide Web is rich with such resources, a number of which are listed in Table 4.

11. SOFTWARE

The analysis of genomic data is heavily dependent on computers, bioinformatics, and Internet technologies. In fact, analytic decisions can often be dictated by the limitations of the available software. To the extent that the investigator wishes to explore multiple analytic options, a familiarity with software applications, such as those shown in Table 5, is imperative *(103)*. There are a variety of commercial and public-domain software applications for each step in the microarray experiment. In an optimal setting, a laboratory information management system (LIMS) will be in place to assist in tracking samples names, sample information, experimental conditions, and protocols. In so doing, the investigator can simplify data reporting at the time of publication as well as ensuring careful data management for the analysis. A variety of image analysis software is available depending on whether cDNA arrays or oligonucleotide arrays are used. Several authors have evaluated the relative strengths and weakness of the methods used by different programs; however, there is no current standard.

The major current focus of software development is in the analysis of microarray data. Microarray data can be analyzed using a range of traditional software from basic spreadsheets such as Microsoft's Excel or more devoted statistical packages such as SAS. In general, however, investigators have preferred dedicated applications with tools focused specifically at microarrays. One of the most flexible of these is the R statistical programming language. R is a programming language in which users can either adopt a wide variety of previously written packages to their specific needs or, if need be, write their own. These multiple applications can be performed without the need to transport data from one format to another, as all are executed within the R environment. Most other software applications offer a fixed number of analytic options without the ability to install or adapt them to the specific needs of the user. Although the user might be able to perform the range of applications desired by transferring data from one application to another, this usually requires reformatting entire datasets. The disadvantage to R is that compared to the stand-alone applications, it is relatively less user friendly.

Table 5 lists a variety of popular software options that perform many of the supervised and unsupervised learning techniques discussed in the previous sections, as well as gene normalization, gene filtering, and data display. Supervised and unsupervised learning algorithms generally divide samples based on lists of genes. For example, the k nearest-neighbors technique chooses a set of genes to define the gene space that best defines clear classes. Lists of genes generated in this way, although more manageable than the thousands of genes originally analyzed in the array experiment, can still prove difficult to interpret. The applications Genmapp, GoSurfer, and Mappbuilder deserve special mention, as they play a role in organizing the long lists of genes generated by clustering and classification algorithms. GenMapp, for example, allows the user to examine genomic data using the Gene Ontology Consortium's GO terms or other user specified functional annotation and provides another potentially useful look at differences between groups.

A final group of applications overlaps with those resources listed in Table 4. These are not stand-alone software but Internet-based tools. Some of these such as the BASE application serve a similar purpose to the programs listed in Table 5, allowing for data viewing or other analysis. More commonly though, these resources are interactive databases and search

Table 5
Software

Application	*Software title*	*Source*
Laboratory Information management systems	MADAM	http://www.tigr.org/software/
Image analysis	ScanAlyze	http://rana.lbl.gov/EisenSoftware.htm
	dChip	http://biosun1.harvard.edu/complab/dchip/
	Spotfinder	http://www.tigr.org/software/
	Microarray Suite[a]	http://www.affymetrix.com
Normalization/analysis/ data display	R Statistical Programming Language	http://www.bioconductor.org/
	MeV	http://www.tigr.org/software/
	Genmapp and Mappbuilder	http://www.genmapp.org/
	Cluster and Treeview	http://rana.lbl.gov/EisenSoftware.htm
	Genecluster	http://www-genome.wi.mit.edu/cancer/
	BASE	http://base.thep.lu.se/
	Microarray Suite[a]	http://www.affymetrix.com
	dChip	http://biosun1.harvard.edu/complab/dchip/
	GoSurfer	http://www.biostat.harvard.edu/complab/gosurfer/

[a]Commercial product.

engines that serve to supplement gene annotation. For example, Affymetrix refers to probes on its arrays by an internal numbering system called the probeset number that can be difficult to use for comparison across different chips. In order to rename the probes in a more standard manner, the user can link Affymetrix probeset number to GenBank or UniGene data via an interactive search on their website.

12. EXAMPLES

Having described the microarray experiment from chip design to analysis, it is a useful exercise to practice the lessons of this chapter by examining two key microarray publications. The first of these, "Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring" by Golub et al. published in 1999, was one of the earliest in the field, yet remains current in terms of its methodology *(6)*. The investigators selected acute leukemia as their model biological system, using only mononuclear cells obtained from bone marrow biopsies, thus avoiding many sample preparation concerns discussed in Section 4. The authors were provided with Affymetrix oligonucleotide arrays, a platform with which the authors at the Whitehead Institute's Center for Genomic Research have worked extensively. Because the array was commercially available, there is little discussion on its specifics characteristics in this publication.

In the initial gene-filtering step, the authors select a subset of informative genes, 1100 out of a possible 6817 genes that correlated with the study parameter of interest, acute leukemia subtype. The authors asked the supervised learning question, How can the subset of genes be used to differentiate the known acute leukemia subtypes of acute myeloid leukemia and acute lymphoid leukemia? They used a classifier called weighted voting, implemented by the software Genecluster, and selected a 50-gene model to make class predictions. Although we have not discussed the weighted voting classifier, we can still evaluate its performance in the validation process. The authors performed the two validation steps we discussed in Section 9.7: internal and external validation. The internal validation was the leave-one-out method, similar to the bootstrap, in which weighted voting classifiers were constructed for datasets that serially excluded a single sample. Each classifier was applied to the single sample that had been excluded from the data used to develop the classifier, and a cumulative error rate was calculated. In this application, the weighted voting method correctly assigned 36/38 leukemias, making no errors but making no call on 2 samples. Subsequently, all samples were used to develop a final weighted voting classifier, which was applied to an independent dataset, again with good performance. In an independent sample, the classifier made no errors in 29/34 samples but made no call in 5.

The authors then ask the unsupervised learning question: If subtypes of leukemia had not been known, could they have discovered them based on strong gene patterns in the data? In this analysis the authors selected the full set of 6817 genes and used the self-organizing maps (SOMs) partitioning method. As we discussed, portioning methods require that the user specify a number of clusters *a priori*—in this case, two. Also recall that unsupervised learning looks for patterns in the data without regard to a named class. Looking only at the data structure with the specification that it be divided into two groups, the SOM method did in fact "discover" the tumor subtypes called A1 and A2, corresponding almost perfectly to the known actue lymphoblastic leukemia (ALL) and acute myoblastic leukemia (AML) subtypes. Once an unsupervised learning method has identified a data cluster, the group and the samples belonging to it can be named—A1 and A2 for example. Once a sample belongs to a named class, supervised learning can be performed so that future samples of that type can be identified. The investigators performed that experiment, building a new classifier (supervised learning experiment) to identify A1 and A2. The new classifier was applied to the independent data, assigning samples to either class A1 or A2 class or uncertain. In almost all cases, A1 status was assigned to ALL and A2 to AML. The example here is contrived in that the "correct" tumor class was previously known. However, the AML–ALL distinction had

not been known, the authors demonstrate that it could have been discovered using this method.

The second publication examines a more concrete clinical problem, where a single histologic tumor type seems to have subtypes based on heterogeneity of clinical behavior *(16)*. The tumor of interest is the diffuse large B-cell lymphoma (DLBCL), where histology alone fails to identify subclasses, yet patients demonstrate markedly different survival. In 2000, Alizadeh et al. reported the cDNA microarray analysis of 96 lymphoma and normal tissue samples. As with the previous example, lymphoid cells avoid many of the difficulties associated with solid tumor samples. The investigators in that study have developed a cDNA array, describing its properties in their introduction. As described in Section 3.2, the cDNA assay requires a reference RNA for which pooled lymphoma cell lines were used.

Lymphomas are divided into several types, of which DLBCL is one subtype, and several lymphoma types were included in the analysis. Although the ability to describe known lymphoma subtypes might best be approached via supervised learning techniques, this was not the stated primary goal of the analysis. The author wished to find previously unknown DLBCL subtypes, a hypothesis that lends itself to more unsupervised learning techniques. Alizadeh et al. applied hierarchical clustering to the entire sample set, which includes a range of samples of known histology. Hierarchical clustering identified unique clusters for all known histologies and suggested two major clusters of DLCBL. The clusters within the histology of DLBCL suggested that one histologic class might be divisible into two distinct gene expression classes. The authors discuss a number of genes that are differentially expressed between the classes, theorizing that one cluster represents a germinal-center-like B-cell tumor and the other an activated-B-cell-like tumor. Finally, the authors examined clinical outcomes according to the cluster assignments for the two DLCBL subtypes, demonstrating a clearly worse prognosis for the activated-B-cell-like tumor. We showed in the leukemia example that once a tumor subtype has been suggested by unsupervised learning, it can be named and supervised learning applied. The investigators chose not to pursue the supervised learning analysis in this case.

The previous two examples are typical of how microarrays analyses are being applied in addressing clinical problems. The methods gathered in this chapter are focused to allow the reader a basic understanding of the relevant techniques at each stage of the analysis. This overview is hopefully specific enough to be concrete, yet acknowledges that an exhaustive review in many cases is not practical. By developing a framework for the generic genomics experiment, including an understanding of the assay, the analytic techniques, and the accompanying databases, the reader can view individual components with flexibility. The individual assays, analytic tools, and databases will evolve; however, the basic principles discussed here will likely prove more durable.

REFERENCES

1. Golub, T. R. Genome-wide views of cancer. *N. Engl. J. Med.* **344:**601–602, 2001.
2. Mohr, S., Leikauf, G. D., Keith, G., and Rihn, B. H. Microarrays as cancer keys: an array of possibilities. *J. Clin. Oncol*. **20:**3165–3175, 2002.
3. Chung, C. H., Bernard, P. S., and Perou, C. M. Molecular portraits and the family tree of cancer. *Nat. Genet.* **32(Suppl.):**533–540, 2002.
4. Bain, B. J. Routine and specialised techniques in the diagnosis of haematological neoplasms. *J. Clin. Pathol*. **48:**501–508, 1995.
5. Perou, C. M., Sorlie, T., Eisen, M. B., et al. Molecular portraits of human breast tumours. *Nature* **406:**747–752, 2000.
6. Golub, T. R., Slonim, D. K., Tamayo, P., et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286:**531–537, 1999.
7. Schena, M., Shalon, D., Davis R. W., and Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270:**467–470, 1995.
8. Ashburner, M., Ball, C. A., Blake, J. A., et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25:**25–29, 2000.
9. Lorentz, C. P., Wieben, E. D., Tefferi, A., Whiteman, D. A, and Dewald, G. W. Primer on medical genomics part I. History of genetics and sequencing of the human genome. *Mayo Clin. Proc*. **77:**773–782, 2002.
10. Stoeckert, C. J., Jr., Causton, H. C., and Ball, C. A. Microarray databases: standards and ontologies. *Nat. Genet.* **32 (Suppl.):**469–473, 2002.
11. Lander, E. S., Linton, L. M., Birren, B., et al. Initial sequencing and analysis of the human genome. *Nature* **409:**860–921, 2001.
12. van de Vijver, M. J., He, Y. D., van't Veer, L. J., et al. A gene–expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347:**1999–2009, 2002.
13. Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O., and Davis, R. W. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci. USA* **93:**10,614–10,619, 1996.
14. Bhattacharjee, A., Richards, W. G., Staunton, J., et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl. Acad. Sci. USA* **98:**13,790–13,795, 2001.
15. Ramaswamy, S., Ross, K. N., Lander, E. S., and Golub, T. R. A molecular signature of metastasis in primary solid tumors. *Nat. Genet.* **33:**49–54, 2003.
16. Alizadeh, A. A., Eisen, M. B., Davis, R. E., et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403:**503–511, 2000.
17. Brown, P. Available at http://cmgm.stanford.edu/pbrown/. 2003.
18. Gershon, D. Microarray technology: an array of opportunities. *Nature* **416:**885–891, 2002.
19. Tefferi, A., Bolander, M. E., Ansell, S. M., Wieben, E. D., and Spelsberg, T. C. Primer on medical genomics. Part III: Microarray experiments and data analysis. *Mayo Clin. Proc.* **77:**927–940, 2002.
20. Cheung, V. G., Morley, M., Aguilar, F., Massimi, A., Kucherlapati, R., and Childs, G. Making and reading microarrays. *Nat. Genet.* **21:**15–19, 1999.
21. Duggan, D. J., Bittner, M., Chen, Y., Meltzer, P., and Trent, J. M. Expression profiling using cDNA microarrays. *Nat. Genet.* **21:**10–14, 1999.
22. Ermolaeva, O., Rastogi, M., Pruitt, K. D., et al. Data management and analysis for gene expression arrays. *Nat. Genet.* **20:**19–23, 1998.
23. Qiagen. *RNeasy Mini Handbook*, 3rd ed., Qiagen, Valencia, CA, 2001.
24. Ramdas, L., Wang, J., Hu, L., Cogdell, D., Taylor, E., and Zhang, W. Comparative evaluation of laser-based microarray scanners. *Biotechniques* **31:**546, 548, 550, passim, 2001.
25. Kohane, I. S., Kho, A. T., and Butte, A. J. *Microarrays for an Integrative Genomics. Computational Molecular Biology*. MIT Press, Cambridge, MA, pp. xviii and 306, 2003.
26. Quackenbush, J. Microarray data normalization and transformation. *Nat. Genet.* **32(Suppl.):**496–501, 2002.
27. King, H. C. and Sinha, A. A. Gene expression profile analysis by DNA microarrays: promise and pitfalls. *JAMA* **286:**2280–2288, 2001.

28. Oleksiak, M. F., Churchill, G. A., and Crawford, D. L. Variation in gene expression within and among natural populations. *Nat. Genet.* **32:**261–266, 2002.
29. Churchill, G. A. Fundamentals of experimental design for cDNA microarrays. *Nat. Genet.* **32(Suppl.):**490–495, 2002.
30. Affymetrix. *GeneChip Arrays for DNA Analysis*, Affymetrix, Santa Clara, CA, 2003.
31. Lipshutz, R. J., Fodor, S. P., Gingeras, T. R., and Lockhart, D. J. High density synthetic oligonucleotide arrays. *Nat. Genet.* **21:**20–24, 1999.
32. Li, C. Available at http://biosun1.harvard.edu/complab/dchip/, 2003.
33. Irizarry, R., Hobbs, B., Collin, F., et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4:**249–264, 2003.
34. Li, C., and Hung Wong, W. Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biol*. **2:**RESEARCH0032, 2001.
35. Affymetrix. *Microarray Suite User's Guide v.5.0.*, Affymetrix, Santa Clara, CA, 2003.
36. Blanchard, A. P. and Hood, L. Hight-density oligonucleotide arrays. *Biosens. Bioelectr*. **11:**687–690, 1996.
37. Singh-Gasson, S., Green, R. D., Yue, Y., et al. Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nat. Biotechnol.* **17:**974–978, 1999.
38. Hughes, T. R., Mao, M., Jones, A. R., et al. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat. Biotechnol.* **19:**342–347, 2001.
39. McGall, G., Labadie, J., Brock, P., Wallraff, G., Nguyen, T., and Hinsberg, W. Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists. *Proc. Natl. Acad. Sci. USA* **93:**13,555–13,560, 1996.
40. Emmert-Buck, M. R., Bonner, R. F., Smith, P. D., et al. Laser capture microdissection. *Science* **274:**998–1001, 1996.
41. Luo, L., Salunga, R. C., Guo, H., et al. Gene expression profiles of laser-captured adjacent neuronal subtypes. *Nat. Med.* **5:**117–122, 1999.
42. Tannapfel, A., Geissler, F., Witzigmann, H., Hauss, J., and Wittekind, C. Analysis of liver allograft rejection related genes using cDNA-microarrays in liver allograft specimen. *Transplant. Proc.* **33:** 3283–3284, 2001.
43. Kitahara, O., Furukawa, Y., Tanaka, T., et al. Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia. *Cancer Res.* **61:**3544–3549, 2001.
44. Curran, S., McKay, J. A., McLeod, H. L., and Murray, G. I. Laser capture microscopy. *Mol. Pathol.* **53:**64–68, 2000.
45. Rubin, M. A. Use of laser capture microdissection, cDNA microarrays, and tissue microarrays in advancing our understanding of prostate cancer. *J. Pathol.* **195:**80–86, 2001.
46. Wang, E., Miller, L. D., Ohnmacht, G. A., Liu, E. T., and Marincola, F. M. High-fidelity mRNA amplification for gene profiling. *Nat. Biotechnol.* **18:**457–459, 2000.
47. Aoyagi, K., Tatsuta, T., Nishigaki, M., et al. A faithful method for PCR-mediated global mRNA amplification and its integration into microarray analysis on laser–captured cells. *Biochem. Biophys. Res. Commun*. **300:**915–920, 2003.
48. Iscove, N. N., Barbara, M., Gu, M., Gibson, M., Modi, C., and Winegarden, N. Representation is faithfully preserved in global cDNA amplified exponentially from sub-picogram quantities of mRNA. *Nat. Biotechnol.* **20:**940–943, 2002.
49. Feldman, A. L., Costouros, N. G., Wang, E., et al. Advantages of mRNA amplification for microarray analysis. *Biotechniques* **33:**906–9112, 914, 2002.
50. Lee, M. L., Kuo, F. C., Whitmore, G. A., and Sklar, J. Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc. Natl. Acad. Sci. USA* **97:**9834–9839, 2000.
51. Kerr, M. K. and Churchill, G. A. Statistical design and the analysis of gene expression microarray data. *Genet. Res.* **77:**123–128, 2001.
52. Brown, P. O. and Botstein, D. Exploring the new world of the genome with DNA microarrays. *Nat. Genet.* **21:**33–37, 1999.
53. Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol. Cell Biol*. **19:**1720–1730, 1999.
54. Chuaqui, R. F., Bonner, R. F., Best, C. J., et al. Post-analysis follow-up and validation of microarray experiments. *Nat. Genet.* **32(Suppl.):**509–514, 2002.
55. Anderson, L. and Seilhamer, J. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **18:**533–537, 1997.
56. Golub, T. R. Genomic approaches to the pathogenesis of hematologic malignancy. Curr. Opin. Hematol. **8:**252–261, 2001.
57. Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. Serial analysis of gene expression. *Science* **270:**484–487, 1995.
58. Patino, W. D., Mian, O. Y., and Hwang, P. M. Serial analysis of gene expression: technical considerations and applications to cardiovascular biology. *Circ. Res.* **91:**565–569, 2002.
59. Polyak, K. and Riggins, G. J. Gene discovery using the serial analysis of gene expression technique: implications for cancer research. *J. Clin. Oncol.* **19:**2948–2958, 2001.
60. Ishii, M., Hashimoto, S., Tsutsumi, S., et al. Direct comparison of GeneChip and SAGE on the quantitative accuracy in transcript profiling analysis. *Genomics* **68:**136–143, 2000.
61. Kallioniemi, A., Kallioniemi, O. P., Sudar, D., et al. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* **258:**818–821, 1992.
62. Pinkel, D., Segraves, R., Sudar, D., et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.* **20:**207–211, 1998.
63. Pollack, J. R., Perou, C. M., Alizadeh, A.A., et al. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* **23:**41–46, 1999.
64. Wang, D. G., Fan, J. B., Siao, C. J., et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280:**1077–1082, 1998.
65. Lindblad-Toh, K., Tanenbaum, D. M., Daly, M. J., et al. Loss-of-heterozygosity analysis of small-cell lung carcinomas using single-nucleotide polymorphism arrays. *Nat. Biotechnol.* **18:**1001–1005, 2000.
66. Packeisen, J., Buerger, H., Krech, R., and Boecker, W. Tissue microarrays: a new approach for quality control in immunohistochemistry. *J. Clin. Pathol.* **55:**613–615, 2002.
67. Camp, R. L., Charette, L. A., and Rimm, D. L. Validation of tissue microarray technology in breast carcinoma. *Lab. Invest.* **80:**1943–1949, 2000.
68. Chee, M., Yang, R., Hubbell, E., et al. Accessing genetic information with high-density DNA arrays. *Science* **274:**610–614, 1996.
69. Hedenfalk, I., Duggan, D., Chen, Y., et al. Gene-expression profiles in hereditary breast cancer. *N. Engl. J. Med.* **344:**539–548, 2001.
70. Mir, K. U. and Southern, E. M. Sequence variation in genes and genomic DNA: methods for large-scale analysis. *Annu. Rev. Genomics Hum. Genet.* **1:**329–360, 2000.
71. Hacia, J. G. and Collins, F. S. Mutational analysis using oligonucleotide microarrays. *J. Med. Genet.* **36:**730–736, 1999.
72. Liotta, L. A. and Petricoin, E. F. Beyond the genome to tissue proteomics. *Breast Cancer Res.* **2:**13–14, 2000.
73. Fields, S. Proteomics. Proteomics in genomeland. *Science* **291:**1221–1224, 2001.
74. Pandey, A. and Mann, M. Proteomics to study genes and genomes. *Nature* **405:**837–846, 2000.
75. Pardanani, A., Wieben, E. D., Spelsberg, T. C., and Tefferi, A. Primer on medical genomics. Part IV: Expression proteomics. *Mayo Clin. Proc.* **77:**1185–1196, 2002.
76. Kepler, T. B., Crosby, L., and Morgan, K. T. Normalization and analysis of DNA microarray data by self-consistency and local regression. *Genome Biol.* **3:**RESEARCH0037, 2002.
77. Zhu, H., Bilgin, M., Bangham, R., et al. Global analysis of protein activities using proteome chips. *Science* **293:**2101–2105, 2001.
78. Siuzdak, G. The emergence of mass spectrometry in biochemical research. *Proc. Natl. Acad. Sci. USA* **91:**11,290–11,297, 1994.

79. Buetow, K. H., Edmonson, M., MacDonald, R., et al. High-throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Proc. Natl. Acad. Sci. USA* **98:**581–584, 2001.
80. Toronen, P., Kolehmainen, M., Wong, G., and Castren, E. Analysis of gene expression data using self-organizing maps. *FEBS Lett.* **451:**142–146, 1999.
81. Rhodes, D. R., Barrette, T. R., Rubin, M. A., Ghosh, D., and Chinnaiyan, A. M. Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Res.* **62:**4427–4433, 2002.
82. van't Veer, L. J., Dai, H., van de Vijver, M. J., et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415:**530–536, 2002.
83. Hoffmann, R., Seidl, T., and Dugas, M. Profound effect of normalization on detection of differentially expressed genes in oligonucleotide microarray data analysis. *Genome Biol.* **3:**RESEARCH0033, 2002.
84. Tseng, G. C., Oh, M. K., Rohlin, L., Liao, J. C., and Wong, W. H. Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Res.* **29:**2549–2557, 2001.
85. Schadt, E. E., Li, C., Ellis, B., and Wong, W. H. Feature extraction and normalization algorithms for high-density oligonucleotide gene expression array data. *J. Cell Biochem.* **37(Suppl.):**120–125, 2001.
86. Dudoit, S., Shaffer, J. P., and Boldrick, J. C. Multiple hypothesis testing in microarray experiments. U. C. Berkeley Division of Biostatistics Working Paper Series, Working Paper 110, pp. 1–52, 2002.
87. Wu, T. D. Analysing gene expression data from DNA microarrays to identify candidate genes. *J. Pathol.* **195:**53–65, 2001.
88. Heyer, L. J., Kruglyak, S., and Yooseph, S. Exploring expression data: identification and analysis of coexpressed genes. *Genome Res.* **9:**1106–1115, 1999.
89. Duda, R. O., Hart, P. E., and Stork, D. G. *Pattern Classification*. Wiley, New York, pp. xx and 654, 2001.
90. Brazma, A. and Vilo, J. Gene expression data analysis. *FEBS Lett.* **480:**17–24, 2000.
91. Sherlock, G. Analysis of large-scale gene expression data. *Curr. Opin. Immunol.* **12:**201–205, 2000.
92. Gasch, A. P. and Eisen, M. B. Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering. *Genome Biol.* **3:**RESEARCH0059, 2002.
93. Yeang, C. H., Ramaswamy, S., Tamayo, P., et al. Molecular classification of multiple tumor types. *Bioinformatics* **17(Suppl. 1):** S316–S322, 2001.
94. Li, L., Weinberg, C. R., Darden, T. A., and Pedersen, L. G. Gene selection for sample classification based on gene expression data: study of sensitivity to choice of parameters of the GA/KNN method. *Bioinformatics* **17:**1131–1142, 2001.
95. Tamames, J., Clark, D., Herrero, J., et al. Bioinformatics methods for the analysis of expression arrays: data clustering and information extraction. *J. Biotechnol.* **98:**269–283, 2002.
96. Berrar, D. P., Downes, C. S., and Dubitzky, W. Multiclass cancer classification using gene expression profiling and probabilistic neural networks. Pacific Symposium on Biocomputing, pp. 5–16, 2003.
97. Furey, T. S., Cristianini, N., Duffy, N., Bednarski, D. W., Schummer, M., and Haussler, D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* **16:**906–914, 2000.
98. Ramaswamy, S., Tamayo, P., Rifkin, R., et al. Multiclass cancer diagnosis using tumor gene expression signatures. *Proc. Natl. Acad. Sci. USA* **98:**15,149–15,154, 2001.
99. Tan, A. C. and Gilbert, D. An empirical comparison of supervised machine learning techniques in bioinformatics, Asian Pacific Bioinformatics Conference, Adelaide, Australia, 2003.
100. Venables, W. N. and Ripley, B. D. *Modern applied Statistics with S. Statistics and Computing.* Springer-Verlag, New York, pp. xi and 495, 2002.
101. Ripley, B. D. *Pattern Recognition and Neural Networks*, Cambridge University Press, Cambridge, pp. xi, and 403, 1996.
102. Brazma, A., Hingamp, P., Quackenbush, J., et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.* **29:**365–371, 2001.
103. Dudoit, S., Gentleman, R. C., and Quackenbush, J. Open source software for the analysis of microarray data. *Biotechniques* Suppl:45–51, 2003.

12 Methods for Analysis of DNA Methylation

ALEXANDER DOBROVIC

1. AN OVERVIEW OF DNA METHYLATION IN HEALTH AND DISEASE

1.1. INTRODUCTION Although assessing DNA methylation has not yet become commonplace in the diagnostic molecular pathology laboratory, many tests involving methylation analysis will become part of routine practice. In this chapter, a practical approach will be taken toward evaluating DNA methylation. The principal question likely to be asked in the pathology laboratory is whether a specific gene or region is methylated in a specific pathological situation; for example, is the *MLH1* gene promoter methylated in a colorectal cancer specimen showing microsatellite instability? This chapter will deal with methods that examine specific sequences for methylation rather than those examining total genomic methylation or those screening for new sites of recurrent methylation.

The various methods used to study methylation each have their particular advantages and disadvantages. The methods chosen should be dictated by the nature of the sample that needs to be analyzed and the information required. As the amount of tissue available to the molecular pathology laboratory is often limited and the DNA is frequently very fragmented, this chapter will be concerned with polymerase chain reaction (PCR)-based methodologies. It will also focus on those methodologies that are used or are likely to be utilized by diagnostic laboratories.

1.2. DNA METHYLATION IN NORMAL CELLS DNA methylation in eukaryotes occurs as a result of the addition of methyl groups to cytosine to form 5-methylcytosine. In humans, DNA methylation occurs almost exclusively at cytosines located within CpG dinucleotides. The CpG dinucleotide (the p denotes physical linkage via a phosphodiester bond) is the unit of DNA methylation. It pairs with a CpG dinucleotide on the opposite DNA strand. This one-to-one correspondence of CpGs enables the replication of methylation patterns immediately following DNA replication during which the methylation of CpGs on the parental strands acts as a template for the corresponding CpGs on the newly synthesized daughter strands.

CpG dinucleotides, which are comparatively rare in DNA, are generally methylated. The observed frequency of CpG dinucleotides in most regions of the DNA is much lower than the expected frequency that would be predicted from the product of the frequencies of the cytosine residues and the guanine residues in those regions *(1)*. This is a result of the accelerated loss of methylated cytosine residues over evolutionary time. When spontaneous cytosine deamination occurs, the resultant uracil is excised by uracil glycosylase in the first step of base excision repair. On the other hand, 5-methylcytosine is deaminated to thymine. The resultant T : G mismatch is recognized by thymine DNA glycosylases, which then excise the thymine. The repair rate is slightly less efficient for 5-methylcytosines than for cytosines probably because uracils, which are not normally part of DNA, are more readily recognized than mismatched thymines. The mutability of 5-methylcytosine underlies much somatic variation and pathogenic mutations in the coding regions of genes *(2,3)*. The CpG dinucleotide becomes either a TpG or a CpA, depending on which strand the deamination has occurred.

CpG islands are cytosine- and guanine- rich segments of the genome with minimal CpG suppression *(1,4)*. They are often associated with the promoter region of genes, and the CpG dinucleotides within the island are generally unmethylated. Methylation of a CpG island in the promoter region of a gene acts to turn off (silence) gene transcription by recruiting histone deacetylases and inducing the formation of inactive chromatin *(5,6)*. In the remainder of this chapter, when we refer to methylation, we will specifically refer to promoter region CpG island methylation.

Methylated DNA influences gene expression via a set of proteins carrying a common methylated CpG- binding domain (MBD). The MBD was first described in the transcriptional repressor MeCP2. Like MeCP2, MBD1, MBD2, and MBD3 are involved in recruiting histone deacetylases to methyl CpG-enriched promoter regions to repress transcription (reviewed in ref. *7*). MBD4 does not affect transcription but is a thymine DNA glycosylase that binds to CpG : TpG mismatches when the cytosine is methylated *(8)*.

Methylation patterns are maintained during replication by the DNA methyltransferase DNMT1, which recognizes the methyl groups on the parental strand and methylates the daughter strand accordingly. The DNA methyltransferases DNMT3A and DNMT3B show *de novo* methylation activity. All three of

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

these enzymes have been shown to be essential for normal embryonic development *(9,10)*.

Two more DNA methyltransferase family members have been described. The role of DNMT2 remains unclear. It is not necessary for either *de novo* or maintenance methylation but might be involved in regulatory pathways as it shows strong binding to DNA *(10,11)*. DNMT3L has been shown to be required for the establishment of maternal methylation imprints at imprinting centers *(12)*. It lacks the catalytic domain common to other DNA methyltransferases but strongly enhances *de novo* methylation by DNMT3A *(13)*.

DNA methylation was postulated to play a central role in the control of gene expression during development *(14,15)*. Although it is no longer considered central, its precise contribution remains poorly understood. Methylation patterns are determined during early embryogenesis by mechanisms that are also poorly understood. Genomewide demethylation after fertilization is followed by *de novo* methylation in pregastrulation *(16)*. The patterns are then faithfully transmitted by DNMT1-mediated maintenance methylation. The instructions for *de novo* methylation must be encoded within the chromatin, as patterns of methylation are highly conserved between individuals *(17)*.

DNA methylation can vary between the two alleles of a gene. DNA methylation is commonly found on the inactive allele when there is differential activity of the two alleles of a gene, as seen in imprinting and X-inactivation. This differential methylation can be used for diagnostic purposes as in the diagnosis of Prader–Willi and Angelman syndromes or the determination of clonality by the X-inactivation method, as will be discussed in Section 3.3.

1.3. DNA METHYLATION IN HUMAN PATHOLOGY

DNA methylation pathways have recently been shown to be directly or indirectly involved in the development of several genetic diseases (reviewed in ref. *18*). Inherited deficiencies affecting the methylation machinery underpin three genetic disorders: ICF, ATR-X, and Rett syndromes. It is of interest that all of these syndromes involve mental retardation as part of their clinical spectrum. It indicates that methylation plays a crucial but as yet unknown role in the development or function of the brain.

Mutations in *DNMT3B* lead to the ICF (immunodeficiency, centromeric region instability, facial abnormalities) syndrome *(19,20)*. Clinically, this disease is characterized by facial dysmorphism, mental retardation, and immune deficiency associated particularly with a decrease in IgA. The heterochromatic regions of chromosomes 1, 9, and 16 are decondensed and show frequent chromatid and chromosome breaks and interchanges *(21,22)*. The centromeric abnormalities arise from lack of methylation of satellite II and III DNA, which is found at the centromeres of chromosomes 1, 9, and 16 *(23)*.

Patients with ATR-X (X-linked α-thalassemia/mental retardation) syndrome have characteristic developmental abnormalities, including severe mental retardation, facial dysmorphism, urogenital abnormalities, and α-thalassaemia. The ATRX gene product is a member of the chromatin remodeling SWI/SNF family and is associated with the pericentromeric heterochromatin. Mutations in the *ATRX* gene also underlie several other X-linked mental retardation syndromes *(24)*. Patients with the ATR-X syndrome show altered methylation of highly repeated sequences in the genome, although why this occurs and the relationship of these changes to the clinical phenotype remains unknown *(25)*.

Mutations in the MBD family gene *MECP2* lead to Rett syndrome, which is characterized by mental retardation, other neurological abnormalities, and mild skeletal abnormalities *(26)*. *MECP2* is X-linked and the disease is normally limited to girls. The mutations are normally lethal in males, although males carrying some of the less deleterious changes can survive past childbirth.

In the Fragile X syndrome, methylation plays a role in the development of the disease, although the methylation machinery is normal. The expansion of a CCG repeat in the 5′ untranslated region past a critical number of copies results in *de novo* methylation of the repeats and inactivation of the adjacent *FMR1* gene *(27,28)*. This same phenomenon is also seen in mental impairment as a result of expansion of CCG repeats at the FRAXE fragile site. The *de novo* methylation of the repeats leads to inactivation of the *FMR2* and *FMR3* genes *(29,30)*.

1.4. DNA METHYLATION CHANGES IN CANCER

Early in cancer development, the distribution of methylation goes awry. Overall methylation is decreased but *de novo* methylation of some promoter-associated CpG islands occurs (reviewed in refs. *31* and *32*). Inactivation by methylation of genes such as tumor suppressor genes, DNA repair genes, and pro-apoptotic genes offers a strong selective advantage to the tumor. Thus, the altered methylation seen during the development of cancer is the origin of much variation that forms the basis for selection of gene expression patterns enhancing the development and progression of malignancy.

A tumor suppressor gene generally requires two "hits" to lose its function; that is, both alleles must be inactivated. Traditionally, one of these hits has been considered to be mutation and the other has been considered to be unmasking of the mutation by physical loss of the second allele (often observed as loss of heterozygosity). Methylation of the tumor-suppressor-gene-promoter region can act as an alternative, either as a first hit or as a second hit. Mutation of one allele can be followed by methylation of the second allele. Methylation of one allele can be followed by physical loss of the second allele or methylation of the second allele.

Methylation has been shown to cause transcriptional silencing of numerous tumor suppressor genes. The pathogenic nature of methylation is illustrated by the observation that for many tumor suppressor genes, methylation of that gene is largely limited to the same type of tumor as that which mutational inactivation occurs; for example, the *RB* gene is methylated in sporadic retinoblastoma *(33–35),* the *VHL* (von Hippel–Lindau) gene is methylated in renal cancer *(36),* and the *MLH1* gene is methylated in colorectal cancer *(37,38)*. The *BRCA1* gene, which is methylated in breast and ovarian cancer, is not inactivated in colon cancer or leukemia *(39,40)*. The gene expression profile of *BRCA1*-methylated breast cancers is closely related to that of cancers that arose from a *BRCA1* mutation, further supporting the direct role of *BRCA1* methylation in the development of breast cancer *(41)*.

Some tumor suppressor genes with promoter region CpG islands do not become methylated in cancer (e.g., the *BRCA2* gene in breast cancer) *(42)*. The reason why some genes with promoter region CpG islands undergo inactivation by methylation and others do not remains unknown. It has been estimated that about 1% of islands can undergo methylation in cancer *(32)*. How much of this methylation is pathogenic remains unknown. In breast cancer, higher-grade tumors have methylation of more CpG islands than the less aggressive lower-grade tumors *(43)*. In colorectal cancer, tumors have been classified as showing a CpG island methylator phenotype (CIMP) if they show methylation of two or more of the five loci; *p16*, *MINT1*, *MINT2*, *MINT31,* and *MLH1 (44)*.

Methylation abnormalities can be reversed by several drugs, notably the inhibitors 5-azacytidine and 5-aza-2′-deoxycytidine. These drugs have been used extensively in vitro to study methylation and are currently undergoing clinical trials as therapeutic agents for several cancers, in particular for the myelodysplastic syndromes, where both drugs have had some success *(45–48)*.

2. PCR-BASED METHODS FOR THE DETECTION OF DNA METHYLATION

2.1. BISULFITE MODIFICATION Any methylation information that is present in the original DNA is lost as soon as the DNA is amplified in vitro. Thus, in order to study methylation using PCR, the DNA must be modified in some way that can be replicated. Sodium bisulfite treatment, which deaminates cytosine to uracil, has become the method of choice for modification. The rate of deamination of 5-methylcytosine to thymine is very much slower than the conversion of cytosine to uracil *(49)*.

Bisulfite modification must be carried out on single-stranded DNA. After this, the cytosines are sulfonated by sodium bisulfite at a low pH; the sulfonated cytosines then readily undergo deamination to sulfonated uracil, which is then desulfonated using high pH.

The use of bisulfite modification prior to PCR treatment enables methylated cytosines in genomic DNA to be directly identified *(50,51)*. The only cytosines remaining derive from methylated CpGs (the sequence 5′ACAT^{m}CGG3′ becomes 5′ATATCGG3′), thus enabling a direct readout of methylated cytosines by sequencing. An important consequence of modification is that the two DNA strands are no longer complementary, with the other strand being converted from 5′C^{m}CGATGT3′ to 5′TCGATGT3′.

Bisulfite modification is robust and a wide variety of conditions have been used. Several solutions (e.g. sodium hydroxide, sodium bisulfite, and hydroquinone) need to be freshly prepared. Many of the important experimental parameters have been described *(52,53)*.

If bisulfite modification is incomplete, accurate analysis of methylation cannot take place. This can be readily observed by sequencing as some non-CpG cytosines will remain as cytosines. A recent review deals with experimental artifacts that inhibit complete bisulfite modification *(54)*. In particular, it is critical to fully denature the DNA before bisulfite modification.

2.2. METHODOLOGIES USED TO STUDY METHYLATION OF BISULFITE- TREATED DNA The gold standard methodology for analysis of methylation is genomic sequencing. However, just as a variety of simpler, more cost-effective methods are often used to replace sequencing in mutation detection, a variety of simpler methods are often used to replace genomic sequencing in methylation detection. Methylation-specific PCR (MSP) is widely used, as it is rapidly performed and can detect relatively low levels of methylation (Section 2.4).

The methodologies can be divided into those that interrogate specific sites for DNA methylation and those that scan a region for DNA methylation. The latter have, in general, been adapted from mutation screening techniques that rely on physical separation according to sequence, such as single-strand conformation analysis, denaturing gradient gel electrophoresis, and denaturing high-performance liquid chromatography. All of the PCR-based techniques can potentially use DNA made from paraffin-embedded formalin–fixed tissues. Several bisulfite modification procedures have been described that deal specifically with DNA from paraffin sections or very low numbers of cells *(55–58)*.

2.3. GENOMIC SEQUENCING The bisulfite-modification-based genomic sequencing methodology *(50,51)* revolutionized the analysis of methylation. Bisulfite-modified DNA is amplified with strand-specific primers framing the region of interest, and the PCR product is then sequenced. The PCR products can be cloned and sequenced to determine the methylation of each cytosine in individual DNA molecules. For tumor samples, this involves sequencing of many clones, as a considerable amount of sequence is often derived from normal tissue. This approach remains an extremely powerful research tool because it allows the detailed study of methylation heterogeneity in any given sample.

The PCR products can be directly sequenced, which gives an average estimation of methylation at each CpG site. Direct sequencing is both faster and considerably less expensive than cloning. However, for heterogeneous samples, it might be difficult to detect low-level methylation. A modification in which only the cytosine and thymine residues are sequenced using fluorescent sequencing and GENESCAN software has been described *(59)*. The degree of methylation is obtained by comparison of the cytosine and thymine peaks.

2.4. METHYLATION-SPECIFIC PCR Methylation-specific PCR (MSP) is the most widely used method for the detection of methylation. It uses primers designed to be specific for methylated, bisulfite-modified DNA *(60)*. Unmethylated sequences are not amplified, although a second pair of primers specific for unmethylated sequences is often designed. MSP is based on the principle that oligonucleotides with a mismatched 3′-residue will not function as primers in the PCR under appropriately stringent conditions. This principle had previously formed the basis of earlier developed single-nucleotide-polymorphism typing methodologies variously known as allele-specific PCR, ARMS (amplification refractory mutation system), and PASA (PCR amplification of specific alleles), which allowed genotyping on the basis of presence or absence of a PCR product *(61–63)*. If a band is seen on a gel after PCR, it is concluded that the sample is methylated or unmethylated according to which primers were used. It is the most sensitive nonquantitative technique available and can detect as low as 0.01% methylation.

A positive MSP signal should only be obtained when both the cytosines recognized by the 3′ end of each of the primers are methylated. This makes MSP an ideal method to screen islands that become heavily methylated. It might be less suited when islands show highly variable methylation, as has been reported for the $p15^{INK4B}$ and *CX26* (connexin 26) genes *(64,65).* MSP could be modified so that only one of the primers recognized a potentially methylatable cytosine, which would allow mapping of one site at a time but MS-SNuPE (Section 2.6) is a better approach to determining single-site methylation.

There are some important limitations of the MSP approach. The first is the susceptibility of MSP to false positives. This can be either a false priming or a sensitivity issue. In the first case, a positive arises even if there is no methylation in the target sequence. If the annealing temperature is too low, amplification can occur across the 3′ mismatch. This type of false positive can be detected by the use of an appropriate control such as a known negative cell line. Raising the annealing temperature and the use of a hot-start methodology or the design of new primers can eliminate this problem. Second, the sensitivity of MSP might lead to false positives because of the amplification of a rare subpopulation of methylated sequences. The tumor sample might be extremely heterogeneous, with only a small proportion of methylated cells. In this case, it would not be correct to call the tumor methylated for that particular gene. Considerable overestimates for the methylation frequency of particular genes have been reported when MSP is used *(66).* There might also be low levels of methylation in the normal somatic tissue *(67)* or incomplete bisulfite conversion.

These limitations arise because MSP is a nonquantitative methodology. It is difficult to tell whether the signal arises from the predominant proportion of cells or from a small subpopulation thereof. Despite these limitations, MSP remains extremely useful, particularly in a preliminary screen of a large number of tumor specimens because of the rapidity with which it is performed.

2.5. BISULFITE PCR FOLLOWED BY RESTRICTION ANALYSIS Several restriction-enzyme-based approaches have been devised to exploit the differences in sequence between methylated and unmethylated DNA after bisulfite modification. The amplified region could include a number of sites that can be analyzed by restriction analysis using enzymes that have a CpG but no other cytosine in their recognition site (such as *Bst*UI, which cuts at CGCG, or *Taq*1, which cuts at TCGA) or by enzymes in which the recognition sequence terminates in a C that is the first base of a CpG residue (such as *Hin*f1, which cuts at GANTC).

*Bst*U1 digestion was first used to show that MSP products were methylated at other CpG sites than those at the 3′ end of the primers *(60).* More commonly, restriction digestion is used with methylation independent PCR (MIP) primers that amplify both methylated and unmethylated strands. After amplification, the DNA is digested with the appropriate restriction enzymes and examined on an agarose or polyacrylamide gel. This approach is sometimes called COBRA (combined bisulfite restriction analysis) *(68,69)*. A variant methodology looks at sites that are newly created following bisulfite modification *(70)*. Because many CpG sites cannot be analyzed, these methods will miss methylation restricted to untested sites.

Restriction analysis can be useful to estimate the degree of methylation, although precise quantitation might be complicated by heteroduplex formation in which strands containing the restriction site anneal with strands not containing the restriction site. This effect becomes more pronounced as the cycle number is increased. In general, these heteroduplexes do not digest with restriction enzymes. A method that eliminates the effect of heteroduplex formation when restriction digestion is used to quantitate alleles has been described *(71)*. Heteroduplex formation might be less likely, however, when there are multiple differences in sequence, as would be seen for methylated and unmethylated PCR products.

2.6. METHYLATION-SENSITIVE SINGLE-NUCLEOTIDE PRIMER EXTENSION ASSAY Methylation-sensitive single-nucleotide primer extension (MS-SNuPE) assay is an adaptation of the single-nucleotide primer extension assay originally introduced to type single-nucleotide polymorphisms *(72)*. It is used to quantitate the relative levels of cytosine and thymine at a single CpG site. Bisulfite-modified DNA is amplified, made single-stranded and then hybridized with an internal primer whose sequence abuts the cytosine of a CpG residue *(73,74)*. Then, DNA polymerase and ^{32}P-dCTP or ^{32}P-dTTP are added in parallel tubes. The internal primer can only be extended if the appropriate deoxynucleotide triphosphate has been added. If the cytosine is methylated, only a deoxcytidine triphosphate can be added, and if it is unmethylated, only a thymidine triphosphate can be added. The reaction is then denatured, electrophoresed on polyacrylamide gels to separate the extended primer from the unincorporated nucleotides, and viewed following autoradiography.

The MS-SNuPE assay is quantitative but seems to have a limited dynamic range. The original methodology only examined the range of 1–100% methylation *(73,74).* This is sufficient for the analysis of most tumor material and could be used to distinguish homozygous from heterozygous methylation in microdissected material. However, it is not suitable where assays that are sensitive over several orders of magnitude are required. Although several CpG sites can be analyzed in a single reaction by using different length primers, this methodology is too labor-intensive in general for screening applications and is mainly applicable to the detailed study of methylation of known genes.

The original method suffered the limitation that radioactivity was used to detect primer extension. In a more recent method, MS-SNuPE products were separated using ion-pair, reverse-phase, high-performance liquid chromatography *(75)*. The methylated and unmethylated CpGs were differentiated and quantified based on the different masses and hydrophobicities of the two extended primer products.

2.7. ENZYMATIC REGIONAL METHYLATION ASSAY Enzymatic regional methylation assay (ERMA) is a quantitative method for determining the methylation density of any DNA region of interest *(76).* This technique is particularly useful for genes like $p15^{INK4B}$ in which methylation density is more important than methylation of particular sites *(64,77)*. Bisulfite modified DNA is amplified with primers containing two dam sites (GATC). The PCR products are purified and incubated with ^{14}C-labeled *S*-adenosyl-methionine (SAM) and

dam methyltransferase, which methylates the adenine residues as a control for the quantity of DNA. A second incubation uses ^{3}H-labeled SAM and SssI methyltransferase to which methylates the CpG cytosines to quantify the CpG sites. The ratio of the $^{3}H/^{14}C$ signals is directly related to the methylation density of the amplified sequence.

2.8. METHYLATION-SENSITIVE SINGLE-STRAND CONFORMATION ANALYSIS Methylation-sensitive single-strand conformation analysis (MS-SSCA) uses single-strand conformation analysis to screen an amplified region of bisulfite-modified DNA for methylation changes *(56,78–81)*. The amplified products are denatured and electrophoresed on a non-denaturing polyacrylamide gel. The sequence differences between unmethylated and methylated sequences lead to the formation of different secondary structures (conformers) with different mobilities. Once the normal (usually unmethylated) pattern is established, any variation would indicate some degree of methylation. Methylated and unmethylated sequences will frequently have multiple-nucleotide differences that will favor the adoption of new conformers. However, there is still the possibility that the conformers might comigrate. Thus, it is advisable to try more than one set of running conditions. Most commonly, two different temperatures—room temperature and 4ºC—are used, although there are several other running conditions that can be useful *(82)*.

Methylation-sensitive SSCA is a convenient method with a very low false-positive rate for screening large numbers of samples for methylation. However, polymorphisms within the region being amplified also might give rise to variant bands. Sequencing is necessary to show whether a variant band is the result of methylation or polymorphism.

MS-SSCA was developed to screen a relatively large number of CpG sites occurring close together, such as those in a CpG island. Whereas all bisulfite methodologies are capable of being used with formalin-fixed paraffin sections, the optimum resolution of MS-SSCA of 150–250 bp is particularly suited to the analysis of DNA derived from this source. All of the CpG sites within a large island can be assessed by developing pairs of primers that will enable the whole island to be amplified in 200–300 bp overlapping fragments. Generally, however, one pair of primers is designed to analyze the sequence flanking the transcriptional start site.

Analysis of tumor material is complex because the specimen can contain substantial amounts of normal cells with normal methylation patterns or there might be intratumor heterogeneity of methylation pattern. MS-SSCA can be used on unfractionated tumor material because it can detect methylation when cells with a methylated DNA sequence comprise less than 10% of the total cells *(56)*. An important advantage of MS-SSCA is that individual bands can be directly sequenced. MS-SSCA can thus be used as the first step for genomic-sequencing studies of methylation. Samples with no identified methylation do not need to be sequenced.

2.9. DNA MELTING ANALYSIS Several methylation screening methods are based on the "melting" properties of DNA in solution. DNA denatures in discrete segments called melting domains as the concentration of a denaturant increases. The melting temperature of a domain is determined by its sequence *(83)*. In denaturing gradient gel electrophoresis (DGGE), PCR products are electrophoresed through a linear gradient of increasing denaturant concentration. When a DNA fragment enters the concentration of denaturant where its lowest temperature domain melts, the molecule changes its structure and migrates more slowly through the gel. This results in separation of different sequences according to their melting properties. The attachment of a GC-rich segment, called a GC clamp, which never denatures at the conditions chosen for the experiment, is necessary to allow for the detection of mutations in the most stable melting domain. DGGE allows detection of virtually all single-base changes in the PCR product.

"Bisulfite-DGGE" uses the same principle to identify methylation *(67)*. Because unmethylated sequences are less cytosine rich than methylated sequences, and partially methylated sequences are intermediate, the fragments separate according to their degree of methylation. This allows clear visualization of heterogeneity of methylation in a way that can be directly related to the density of methylation. By contrast, SSCA will also allow visualization of methylation heterogeneity, but this is not related to methylation density. The DGGE approach is an unusually powerful one but has been adopted by few laboratories.

The simplest methodology based on melting analysis monitors the fluorescence of the double- stranded DNA-binding dye SYBR Green 1 as the double-stranded PCR product is slowly denatured *(84)*. This is readily performed on all real-time PCR machines. The analysis is done in the same tube or well as the PCR, which allows high throughput and reduces problems resulting from PCR product cross-contamination. Standard primers can be used but analysis might be facilitated by designing primers with GC clamps similar to those used in DGGE *(83)*.

Methylation screening of bisulfite-modified PCR products using denaturing high-performance liquid chromatography (DHPLC) is only just beginning to be used but is likely to become an important screening methodology. As with SSCA and DGGE, DHPLC screening does not distinguish CpG methylation and single-nucleotide polymorphisms, and sequencing needs to be used to distinguish between them.

The optimum temperature for a DHPLC run can be predicted using the sequence of the fully methylated product. The temperature then needs to be verified so that tight peaks without much spread are obtained. The retention time of the peak should correlate with methylation status, because the more unmethylated the target is, the less GC rich the PCR product is and the lower the retention time is. DHPLC has been used to differentiate the methylated and unmethylated alleles of imprinted loci *(85)* and MLH1 promoter region methylation *(86)*.

2.10. REAL-TIME PCR METHODOLOGIES The major disadvantage of MSP is that although it is the most sensitive technique for detecting methylation, it is not quantitative. It can, however, be adapted to a real-time PCR approach that will allow quantitation. Importantly, quantitation allows discrimination of signal from low-level background, which might occur in normal tissue. Real-time PCR analysis also requires no further manipulation after the PCR step, which allows high-throughput analysis and eliminates problems resulting from cross-contamination of PCR products.

In the MethyLight methodology *(87,88)*, MSP is combined with the TaqMan methodology, which uses a fluorescent-labeled probe to monitor amplification. Bisulfite-modified DNA is amplified using MSP primers flanking an oligonucleotide probe containing a 5′ fluorescent reporter and a 3′ quencher. The 5′ to 3′ nuclease activity of *Taq* DNA polymerase cleaves the probe, separating the fluorescent reporter from the quencher each time an amplification occurs. The intensity of the resultant fluorescent signal is proportional to the amount of PCR product generated, thus allowing quantification of the PCR reaction.

MethyLight has been called a quantitative MSP assay. However, the most commonly used TaqMan probes contain several CpG sites in their sequence. The effect of the CpG containing probe is to make MethylLight more specific for hypermethylation. However, this might make MethyLight data complex to interpret. According to its design and the PCR conditions used, the TaqMan probe might bind only if all the probe CpG sites are also methylated or might bind when only some of the sites are methylated.

MethyLight is capable of detecting methylated alleles in the presence of a 10,000-fold excess of unmethylated alleles *(87,88)*. Methylation of candidate genes needs to be scored relative to standard curves constructed from dilutions of DNA prepared from a cell line that is methylated for the gene in question. These dilutions can be made in DNA prepared from a cell line that is correspondingly unmethylated.

Sequences that have been incompletely converted during bisulfite treatment might be coamplified during MethyLight, resulting in an overestimation of the level of DNA methylation. In ConLight-MSP, an additional fluorescent probe directed against unconverted DNA is used in the same reaction as the specific TaqMan probe *(89)*.

There are many potential variations involving the confluence of methylation analysis and real-time methodology. For example, nonselective amplification could be used with probes for both methylated and unmethylated DNA in the same tube as a type of allelic discrimination assay. However, the results might not be all that easy to interpret where there is heterogeneity of methylation patterns. The simplest real-time approach would be to convert an established MSP protocol to a quantitative protocol using SYBR Green in the reaction mix. It is essential that the MSP reaction is highly specific, as SYBR Green-based methods cannot readily distinguish between specific and nonspecific reactions.

2.11. PROFILING AND ARRAYS It would be clearly advantageous to analyze a cancer specimen for the methylation of dozens if not hundreds or thousands of CpG islands at the same time. Restriction landmark genomic sequencing is one such methodology *(90,91)* but is technically difficult and is more suited to the research laboratory at its current stage of development. Differential methylation hybridization (DMH) can be used to study genomewide methylation patterns in CpG islands in DNA from cancer cell lines, tumor samples, and normal tissue, but it is also more suitable for gene discovery because the methodology is technically complex *(92)*. Currently, the most promising array-based approaches are oligonucleotide based. Bisulfite-treated DNA samples are amplified and hybridized to arrayed oligonucleotides that are specific for methylated and unmethylated sequences at chosen CpG islands *(93,94)*.

2.12. PRIMER DESIGN Primer design is a critical aspect of PCR-based analysis of bisulfite-modified DNA. Primers can be designed to analyze any region but are most often designed so that the amplified region overlaps the transcriptional start site and consists mainly of sequence prior to the start site. The DNA strands are no longer complementary after bisulfite treatment. Primers are generally designed to amplify from the modified sense strand, but if it is difficult to design effective primers, the modified antisense strand should be examined.

For a new sequence, identifying the promoter region CpG island is the crucial first part of the primer design process. The transcriptional start region of the gene must first be located. Many cDNA sequences include very little 5′ untranslated region sequence. The most recent 5′ cDNA sequence available should be used to Blast search GenBank (http://www.ncbi.nlm.nih.gov/blast) to identify mRNAs or expressed sequence tags that have extra 5′ sequence. Using the TRASER website facilitates this process (http://genome-www6.stanford.edu/cgi-bin/Traser/traser). The most 5′ sequence identified is then used to identify genomic sequence that is likely to be at the transcriptional start site. This genomic sequence can then be examined for the presence of a CpG island. CpG islands can be identified by visually identifying clusters of CpGs as outlined next or using the Web-based program CpG Island Searcher (http://www.uscnorris.com/cpgislands/).

For MSP, regions that contain frequent CpG sites are chosen and many investigators try to fit as many CpG sequences into the 3′ end of the primer as possible. It is desirable to have the 3′ end corresponding to a cytosine of a CpG dinucleotide, but this is not universally done. MSP is best suited for screening genes where the methylation patterns are already known and the 3′ ends can be placed at the most frequently methylated sites.

Quite different requirements are necessary for MIP primers that amplify all converted sequences prior to analysis, as is required for genomic sequencing or MS-SSCA. MIP primers are designed to amplify both bisulfite-modified methylated and unmethylated DNA by either avoiding CpG sites or including as few as possible and placing them as far as possible to the 5′ end of the primer. The primer sequence at the cytosine of a CpG residue can either be (1) C (G on the antisense primer), which biases amplification in favor of methylated sequences, or (2) a C/T degeneracy (G/A on the antisense primer), or (3) an inosine. Because methylated sequences can have an amplification disadvantage, the first option is often acceptable, particularly when techniques screening for methylation are used.

Careful primer choice is important to minimize the amplification of incompletely bisulfite-modified DNA. Primers are chosen that have a T derived from a non-CpG cytosine at or as near as possible to the 3′ end of each primer. A website for designing methylation primers (http://itsa.ucsf.edu/~urolab/methprimer/index1.html) has recently become available *(95)*. Choosing primers that meet the constraints of CpG placement and modified T placement can also be done by visual scanning

of the modified sequence that is facilitated by most word processing programs.

The following approach has proven to be effective for visual scanning for suitable primers. The sense or antisense sequence is converted to lowercase and all spaces and carriage returns are removed. The font is changed to a proportional font such as Courier and the number of characters is adjusted to 50 or 100 per line. The color of the Cs and Gs is changed to green. The color of the CG dinucleotide is changed to red. The colors chosen are arbitrary, but the above combination allows ready visualization of the CpG island both in terms of high G + C content and CpG dinucleotide density. The sequence is copied twice. The first sequence is denoted the (bisulfite) modified unmethylated sequence and all lowercase C's are changed to an uppercase T. The second sequence is denoted the modified methylated sequence and all C's are changed to an uppercase T except for the C's in CpGs. This makes scanning for potential primers very easy because the CpGs are readily visible and the modified C's are readily seen as uppercase T's. This approach can be automated by making a word processing macro *(96)*.

It is often preferable to use primers with high melting temperatures to help with specificity, particularly because the DNA sequence is less complex after bisulfite modification because of the depletion of cytosines. The last five 3′ nucleotides preferably should contain two or three (but not more) G's (C's on the antisense strand) to stabilize the 3′ end of the primer. We use the Web program Oligonucleotide Properties Calculator to check the melting temperatures (http://www.basic.nwu.edu/biotools/oligocalc. html). We find that the most reliable calculation to set the initial annealing temperatures is 5ºC below the salt-adjusted melting temperature. The primers can be further checked using the Macintosh freeware program Amplify (version 1.2) to identify nonspecific primer binding within the region of interest as well as potential primer dimers. We also use Amplify to ensure that the chosen primers do not amplify the unmodified sequence and that the modified methylated and modified unmethylated are both amplified.

The use of nesting is not recommended because of the increased potential for carryover of PCR amplicons. It is unnecessary, in most cases, with primers designed to the above criteria but might be necessary when working from micro-dissected material in order to generate sufficient PCR product.

3. DIAGNOSTIC APPLICATIONS OF DNA METHYLATION

3.1. DNA METHYLATION AS A MARKER OF MALIGNANT DISEASE Tumors show characteristic profiles of methylated markers that are specific for given types and subtypes of cancer *(32,97)*. Thus, there is considerable potential for using methylation as a tumor-specific marker *(98)*. Methylation markers are likely to be the best DNA-based markers because there are no useful high-frequency, single-site mutations known for most cancers. With the right choice of high-frequency methylated genes, tumor cells or DNA can be identified without knowledge of which genes are methylated in the primary tumor.

Currently, the best example of a tumor-specific methylation marker is the *GSTP1* gene, which is methylated in 95% of prostate cancers but not in benign prostatic hyperplasia *(99)*. For other cancers, markers showing this degree of specificity remain to be identified. An alternative approach is to use a small panel of recurrently methylated markers rather than a single marker. By assembling a panel of four to five loci, it might be possible to cover close to 100% of tumors. It is critical to verify that the markers are generally not methylated to any significant degree in normal tissue.

Methylation markers can be used for early detection of cancer or for monitoring response to treatment. Sensitive techniques can be used to detect methylation in biological fluids such as (1) blood for the detection of disseminated carcinoma cells or tumor-derived DNA, or (2) various lumenal fluids such as urine for the detection of cells from bladder, kidney, and prostate cancers, or (3) washings such as bronchioalveolar and ductal lavages for the detection of early malignant cells in lung and breast cancer, respectively.

The detection of methylated sequences in DNA derived from the plasma or serum is a potentially important early detection and monitoring method for cancer *(100)*. DNA alterations have been detected in patients with small or even *in situ* lesions, indicating that tumor DNA is shed into circulation early in the disease. This DNA probably derives from necrosis and apoptosis of the tumor cells. In one study, plasma DNA concentration in normal individuals averaged less than 4 ng/mL, whereas the plasma DNA concentration in breast cancer patients ranged from 20 to 360 ng/mL *(101)*.

The utility of methylation as a tumor marker has been demonstrated by numerous studies, including the following examples. Methylation of the $p16^{INK4A}$ tumor suppressor gene, which was shown to be an early change in lung cancer, was detected in the sputum of 3/7 patients with squamous cell carcinoma and 5/26 high-risk cancer-free individuals *(102)*. In patients with resectable non-small-cell lung cancer, *p16* methylation was detected in the in bronchioalveolar lavage fluid of 12/19 patients whose tumor showed *p16* methylation *(103)*. In patients with prostate cancer , *GSTP1* promoter hypermethylation was detected in 72% of plasma or serum samples, 50% of ejaculates, and 36% of urine samples (after prostate massage) *(104)*. In ductal lavage specimens, methylated alleles of the cyclin D2 (*CCND2*) and *RARβ2* genes were detected in fluid from patients with endoscopically detected carcinomas and ductal carcinoma *in situ* but rarely in fluid from healthy ducts *(105)*.

3.2. DETERMINATION OF CLONALITY The clonality of tumors in females can be determined using methods that utilize methylation differences between the active and inactive X chromosomes *(106)*. The X chromosomes are genetically distinguished using single-nucleotide polymorphisms *(107)* or by variation in copy number of a trinucleotide repeat of the androgen receptor locus *(108)*. The technique uses a methylation-sensitive restriction enzyme to determine which allele of the polymorphic locus on the X chromosome is methylated. The unmethylated allele is digested and, consequently, is not amplifiable *(107)*. Clonal cell populations, which all have the same allele of the X chromosome inactivated, will show preferential loss of one allele after restriction digestion and PCR amplification. One problem in using the clonality determination

approach is that there might be significant variation in the balance or skewing of X chromosome inactivation in normal individuals. Appropriate normal tissue controls from the same individuals should always be employed as a control *(109)*.

3.3. DIAGNOSIS OF IMPRINTING DISORDERS The majority of Prader–Willi syndrome and Angelman syndrome cases have lesions that effectively remove one of the copies of a region in chromosome 15q11-q13 (e.g., deletions or uniparental disomy). Prader–Willi syndrome arises from loss of the paternally derived copy of this region, whereas Angelman syndrome arises from loss of the maternally derived copy of this region. Both disorders are characterized at the molecular level by abnormal methylation of imprinted genes at 15q11-q13, including the small nuclear ribonucleoprotein N gene (*SNRPN*). The promoter region CpG island of the *SNRPN* gene is heavily methylated in the maternally derived allele and unmethylated in the paternally derived allele. In patients with Prader–Willi syndrome, only the methylated allele is present, whereas in patients with Angelman syndrome, only the unmethylated allele is present. Several MSP-based assays have been designed to evaluate the methylation status of the SNRPN CpG island for rapid diagnosis of these two syndromes. Bisulfite-modified DNA from patients with Prader–Willi syndrome only amplify with the methylation-specific pair, whereas modified DNA from patients with Angelman syndrome only amplify with the primers specific for the unmethylated sequence. Modified DNA from normal individuals amplifies with both primer pairs *(110–112)*.

More recently, DHPLC-based approaches have been used to differentiate the two alleles. Both the methylated and unmethylated alleles are amplified at the same time and separated by DHPLC. The sensitivity and semiquantitative properties of DHPLC are able to detect mosaicism more readily than MSP-based approaches. The DHPLC approach has been also applied to other imprinted regions where a methylation difference is also present (e.g., in the Beckwith–Wiedemann syndrome) *(113,114)*.

3.4. DNA METHYLATION AS A PREDICTIVE MARKER Whereas *de novo* methylation of the promoter regions of DNA repair genes can lead to accelerated carcinogenesis by increasing the mutation rate, this loss of DNA repair capacity can also be the Achille's heel for the tumor during chemotherapy and or radiotherapy. Many therapies targeting cancer cells are effective because they cause DNA damage. Knowledge of the affected pathways can thus lead to rational choice of therapeutic agent and allow the prediction of "responders" and "nonresponders."

A compelling example is methylation of the O^6 methylguanine DNA methyltransferase (*MGMT*) gene, which removes small alkyl groups from the O^6 position of guanine. *MGMT* is methylated in a variety of cancers *(115)*. In gliomas and diffuse large B-cell lymphomas, *MGMT* promoter methylation has been shown to be associated with response to chemotherapy with alkylating agents, which are now much more toxic to the deficient cells *(116,117)*. Thus, loss of activity of MGMT, which initially favors tumor progression, is now responsible for the tumor's exquisite sensitivity to alkylating agents.

Methylation of *MLH1*, which leads to deficient mismatch repair in sporadic colon tumors, is another example of how knowledge of the lesion can suggest specific therapy. Mismatch repair (MMR) acts to recognize and process not only single-basepair mismatches and insertion–deletion loops that occur during DNA replication but also DNA adducts such as those resulting from treatment with cancer chemotherapy agents *(118)*. However, whereas base mismatches and insertion–deletion loops are repaired by MMR, MMR-mediated recognition and processing of chemotherapy-induced adducts in DNA results in apoptosis.

The inability of MMR-deficient cells to recognize chemotherapy-induced adducts in DNA leads to resistance to alkylating agents. Interestingly, MMR-positive tumors are more sensitive to standard chemotherapies for colorectal cancer using 5-fluorouracil *(119,120)*. This validates the standard chemotherapy for colon cancer in this group of tumors but suggests that the distinct group with *MGMT* methylation *(121)* might benefit from therapy with alkylating agents.

3.5. DNA METHYLATION AND THE EVALUATION OF HNPCC A particular problem in the study of hereditary nonpolyposis colorectal cancer (HNPCC) is deciding which patients to screen for mutations in one or more of the *MLH1*, *MSH2*, and *MSH6* genes. Pedigrees first need to meet strict criteria such as the Amsterdam criteria to establish a high likelihood of hereditary cancer. A tumor from an affected member of the pedigree is then screened by examination of a panel of mononucleotide and dinucleotide repeats to determine microsatellite instability (MSI) and by immunohistochemistry for the MLH1, MSH2, and MSH6 enzymes. If the tumor is scored as MSI high and/or immunohistochemistry for one or more enzymes is negative, mutation screening is undertaken.

The *MLH1* locus undergoes frequent methylation in sporadic colorectal tumors *(37,38)*. Both *MLH1*-methylated tumors and tumors arising from patients with a germline *MLH1* mutation have the same phenotype: MSI high and negative immunohistochemistry for the MLH1 protein. However, the identification of *MLH1* methylation in a tumor does not disprove its hereditary origin, as methylation can occur as the second event in a patient with a germline mutation. The absence of *MLH1* methylation in tumors of this phenotype is associated with a very strong likelihood of there being a germline mutation. The decision whether to proceed with germline testing should be based on all the known clinical and pathological information *(122)*.

The second case when methylation status can be used in the context of HNPCC screening is that of MSI low tumors. These present somewhat of a diagnostic dilemma because it has been suggested that many tumors identified by the National Cancer Institute panel of five markers *(123)* as MSI high (defined as those tumors that show MSI for more than one marker) are, in fact, MSI low (defined as those tumors that show MSI for only one marker). Certain dinucleotide and tetranucleotide markers show instability at higher than expected frequencies in non-MSI high cancers (e.g., *D2S123*, *MYCL*, and *D17S250)*. The same markers have been recommended for use in the NCI panel. Should two of these markers be mutated in the same cancer, then that cancer can be classified as MSI high. However, such cancers do not show immunohistochemical loss of DNA

MMR proteins, instability in mononucleotide markers, or the clinical and pathological features that have been associated with MSI high cancers *(122)*. Therefore, these cancers should probably be included with MSI-low and not MSI-high cancers. It is considered that in some cases that methylation of the *MGMT* locus can give rise to the MSI-low phenotype *(121)*. In this case, identification of *MGMT* methylation in a MSI-low tumor might confirm the decision not to go ahead with mutation screening.

3.6. QUALITY CONTROL ISSUES All of the quality control issues that apply to PCR-based tests are applicable to the methods described here. As with all PCR, utmost care must be taken to ensure that there is no carryover of PCR amplicons into the PCR setup areas. In addition, a sample of unmodified DNA should always be run to ensure that no amplification is observed under the conditions being used. This is particularly important to avoid false positives when MSP is being performed. The appropriate methylated and unmethylated controls should be used. It is convenient to use cell lines with known methylation status for this. If no positive control cell line is available, Sss1 methylase can be used to methylate all CpG residues in any available DNA. Normal human DNA is preferable to human cancer cell line DNA as a substrate for Sss1 methylation, as there is always the possibility that a cell line contains homozygous deletions for the regions of interest.

A second set of issues arises when quantitation is important. Preferential amplification of either methylated or unmethylated sequences (PCR bias) could occur *(124)*. Usually, methylated sequences amplify less readily than unmethylated sequences, because the higher C+G content of the PCR products leads to their more inefficient denaturation. Mixing known ratios of methylated to unmethylated targets to construct a standard curve can be used to both assess and to compensate for this problem.

4. THE FUTURE

It is always difficult to forecast the future, but several predictions can be made for the next few years. First, real-time PCR will become widely established in the analysis of methylation. Real-time PCR allows faster analysis and higher throughput because there are no further manipulations after the PCR step. However, the principal advantage is that molecular detection of low levels of disease benefits from the absence of post sample processing that real-time PCR analysis allows. Although stringent precautions are taken to eliminate cross-contamination, not having to remove the PCR products from the tube for analysis will eliminate PCR product contamination problems.

High-throughput MethyLight-based methods allowing the analysis of methylation at hundreds of sites at a time will challenge DNA methylation arrays as a method to analyze methylation at multiple sites, particularly because arrays are unlikely to be quantitative over a similar dynamic range. The information gained by multilocus methylation profiling will complement expression profiling as a diagnostic and prognostic tool. Finally, the ultimate methylation analysis method might be based on mass spectrometry that can be done directly from DNA without bisulfite modification.

ACKNOWLEDGMENTS

First, I acknowledge my fellow researchers in the exciting field of DNA methylation and human disease, both those whom I have cited and those whom I have omitted. I would like to thank Denise O'Keefe, Tina Bianco, Susan Clark, Jeremy Jass, Kathie Friend, Nick Wickham, Ted Bradley, Maurice Loughrey, Damian Hussey, Michael Krypoy, and Ella Thompson for reading one or more drafts and/or for their helpful comments. Finally, I would like to thank my wife, JoAnn, for her support while this chapter was being written.

REFERENCES

1. Gardiner-Garden, M. and Frommer, M. CpG islands in vertebrate genomes. *J. Mol. Biol*. **196:**261–282, 1987.
2. Cooper, D. N. and Krawczak, M. Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. *Hum. Genet*. **83:** 181–188, 1989.
3. Krawczak, M., Ball, E. V., and Cooper, D. N. Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes. *Am. J. Hum. Genet*. **63:**474–488, 1998.
4. Bird, A. P. CpG-rich islands and the function of DNA methylation. *Nature* **321:**209–213, 1986.
5. Nan, X., Ng, H. H., Johnson, C. A., et al. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393:**386–389, 1998.
6. Jones, P. L., Veenstra, G. J., Wade, P. A., et al. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet*. **19:**187–191, 1998.
7. Ballestar, E. and Wolffe, A. P. Methyl-CpG-binding proteins. Targeting specific gene repression. *Eur. J. Biochem*. **268:**1–6, 2001.
8. Hendrich, B., Hardeland, U., Ng, H. H., Jiricny, J., and Bird, A. The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature* **401:**301–304, 1991.
9. Li, E., Bestor, T. H., and Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69:**915–926, 1992.
10. Okano, M., Bell, D. W., Haber, D. A., and Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99:**247–257, 1991.
11. Dong, A., Yoder, J. A., Zhang, X., Zhou, L., Bestor, T. H., and Cheng, X. Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA. *Nucleic Acids Res*. **29:**439–448, 2001.
12. Hata, K., Okano, M., Lei, H., and Li, E. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* **129:**1983–1993, 2002.
13. Chedin, F., Lieber, M. R., and Hsieh, C. L. The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proc. Natl. Acad. Sci. USA* **99:**16,916–16,921, 2002.
14. Holliday, R. and Pugh, J. E. DNA modification mechanisms and gene activity during development. *Science* **187:**226–232, 1975.
15. Riggs, A. D. X inactivation, differentiation and DNA methylation. *Cytogenet. Cell. Genet*. **14:**9–25, 1975.
16. Razin, A. and Shemer, R. DNA methylation in early development. *Hum. Mol. Genet*. **4:**1751–1755, 1995.
17. Behn-Krappa, A., Holker, I., Sandaradura de Silva, U., and Doerfler, W. Patterns of DNA methylation are indistinguishable in different individuals over a wide range of human DNA sequences. *Genomics* **11:**1–7, 1991.
18. Robertson, K. D. and Wolffe, A. P. DNA methylation in health and disease. *Nat. Rev. Genet*. **1:** 11–19, 2000.
19. Xu, G. L., Bestor, T. H., Bourc'his, D., et al. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* **402:**187–191, 1999.
20. Hansen, R. S., Wijmenga, C., Luo, P., The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proc. Natl. Acad. Sci. USA* **96:**14,412–14,417, 1999.

21. Fryns, J. P., Azou, M., Jaeken, J., Eggermont, E., Pedersen, J. C., and Van den Berghe, H. Centromeric instability of chromosomes 1, 9 and 16 associated with combined immunodeficiency. *Hum. Genet*. **57:**108–110, 1981.
22. Haas, O. Centromeric heterochromatin instability of chromosomes 1, 9 and 16 in variable immunodeficiency syndrome—a virus-induced phenomenon? *Hum. Genet.* **85:**244–246, 1990.
23. Ehrlich, M. DNA hypomethylation, cancer, the immunodeficiency, centromeric region instability, facial anomalies syndrome and chromosomal rearrangements. *J. Nutr*. **132:**2424S–2429S, 2002.
24. Gibbons, R. J. and Higgs, D. R. Molecular-clinical spectrum of the ATR-X syndrome. *Am. J. Med. Genet*. **97:**204–212, 2002.
25. Gibbons, R. J., McDowell, T. L., Raman, S., et al. Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. *Nat. Genet*. **24:**368–371, 2000.
26. Shahbazian, M. D. and Zoghbi, H. Y. Rett syndrome and MeCP2: linking epigenetics and neuronal function. *Am. J. Hum. Genet*. **71:**1259–1272, 2002.
27. Oberle, I., Rousseau, F., Heitz, D., et al. Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science* **252:**1097–1102, 1991.
28. Kremer, E. J., Pritchard, M., Lynch, M., et al. Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. *Science* **252:**1711–1714, 1991.
29. Gecz, J. The FMR2 gene, FRAXE and non-specific X-linked mental retardation: clinical and molecular aspects. *Ann. Hum. Genet*. **64:**95–106, 2000.
30. Gecz, J. FMR3 is a novel gene associated with FRAXE CpG island and transcriptionally silent in FRAXE full mutations. *J. Med. Genet*. **37:**782–784, 2000.
31. Jones, P. A. and Baylin, S. B. The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet*. **3:**15–428, 2002.
32. Costello, J. F., Fruhwald, M. C., Smiraglia, D. J., et al. Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat. Genet*. **24:**132–138, 2000.
33. Greger, V., Passarge, E., Hopping, W., Messmer, E., and Horsthemke, B. Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. *Hum. Genet*. **83:**155–158, 1989.
34. Sakai, T., Toguchida, J., Ohtani, N., Yandell, D. W., Rapaport, J. M., and Dryja, T. P. Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene. *Am. J. Hum. Genet*. **48:**880–888, 1991.
35. Stirzaker, C., Millar, D. S., Paul, C. L., et al. Extensive DNA methylation spanning the Rb promoter in retinoblastoma tumors. *Cancer Res*. **57:**2229–2237, 1997.
36. Herman, J. G., Latif, F., Weng, Y., et al. Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc. Natl. Acad. Sci. USA*. **91:**9700–9704, 1994.
37. Kane, M. F., Loda, M., Gaida, G. M., et al. Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. *Cancer Res*. **57:**808–811, 1997.
38. Herman, J. G., Umar, A., Polyak, K., et al. Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc. Natl. Acad. Sci. USA* **95:**6870–6875, 1998.
39. Bianco, T., Chenevix-Trench, G., Walsh, D. C., Cooper, J. E. , and Dobrovic, A. Tumour-specific distribution of BRCA1 promoter region methylation supports a pathogenetic role in breast and ovarian cancer. *Carcinogenesis* **21:**147–151, 2000.
40. Esteller, M., Silva, J. M., Dominguez, G., et al. Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *J. Natl. Cancer Inst*. **92:**564–569, 2000.
41. Hedenfalk, I., Duggan, D., Chen, Y., et al. Gene-expression profiles in hereditary breast cancer. *N. Engl. J. Med*. **344:**539–548, 2001.
42. Collins, N., Wooster, R., and Stratton, M. R. Absence of methylation of CpG dinucleotides within the promoter of the breast cancer susceptibility gene BRCA2 in normal tissues and in breast and ovarian cancers. *Br. J. Cancer* **76:**1150–1156, 1997.
43. Yan, P. S., Chen, C. M., Shi, H., et al. Dissecting complex epigenetic alterations in breast cancer using CpG island microarrays. *Cancer Res*. **61:**8375–8380, 2001.
44. Toyota, M., Ahuja, N., Ohe-Toyota, M., et al. CpG island methylator phenotype in colorectal cancer. *Proc. Natl. Acad. Sci. USA* **96:**8681–8686, 2001.
45. Santini, V., Kantarjian, H. M., and Issa, J. P. Changes in DNA methylation in neoplasia: pathophysiology and therapeutic implications. *Ann. Intern. Med*. **134:**573–586, 2001.
46. List, A. F. New approaches to the treatment of myelodysplasia. *Oncologist* **7**(Suppl . 1):39–49, 2002.
47. Daskalakis, M., Nguyen, T. T., Nguyen, C., et al. Demethylation of a hypermethylated P15/INK4B gene in patients with myelodysplastic syndrome by 5-Aza-2′-deoxycytidine (decitabine) treatment. *Blood* **100:**2957–2964, 2002.
48. Leone, G., Teofili, L., Voso, M. T., and Lubbert, M. DNA methylation and demethylating drugs in myelodysplastic syndromes and secondary leukemias. *Haematologica* **87:**1324–1341, 2002.
49. Wang, R. Y., Gehrke, C. W., and Ehrlich, M. Comparison of bisulfite modification of 5-methyldeoxycytidine and deoxycytidine residues. *Nucleic Acids Res*. **8:**4777–4790, 1980.
50. Frommer, M., Mcdonald, L. E., Millar, D. S., et al. A Genomic Sequencing Protocol That Yields a Positive Display of 5-Methylcytosine Residues in Individual DNA Strands. *Proc. Natl. Acad.Sci. USA* **89:**1827–1831, 1992.
51. Clark, S. J., Harrison, J., Paul, C. L., and Frommer, M. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res*. **22:**2990–2997, 1994.
52. Clark, S. J. and Frommer, M. F. Deamination with NaHSO3 in DNA methylation studies, in *DNA and Nucleoprotein Structure InVivo*, Saluz, H. and Wiebauer, K., eds., Biomedical Publishers. pp. 123–132, 1995.
53. Grunau, C., Clark, S. J., and Rosenthal, A. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Res*. **29:**E65,5, 2001.
54. Warnecke, P. M., Stirzaker, C., Song, J., Grunau, C., Melki, J. R., and Clark, S. J. Identification and resolution of artifacts in bisulfite sequencing. *Methods*. **27:**101–107, 2002.
55. McDonald, L. E. and Kay, G. F. Methylation analysis using bisulfite genomic sequencing: application to small numbers of intact cells. *Biotechniques* **22:**272–274, 1997.
56. Bian, Y. S., Yan, P., Osterheld, M. C., Fontolliet, C., and Benhattar, J. Promoter methylation analysis on microdissected paraffin-embedded tissues using bisulfite treatment and PCR-SSCP. *Biotechniques* **30:**66–72, 2001.
57. Tan, L. W. and Dobrovic, A. Methylation analysis of formalin-fixed, paraffin-embedded sections using a nontoxic DNA extraction protocol. *Biotechniques* **31:**1354, 1356–1357, 2001.
58. Millar, D. S., Warnecke, P. M., Melki, J. R., and Clark, S. J. Methylation sequencing from limiting DNA: embryonic, fixed and microdissected cells. *Methods* **27:**108–113, 2002.
59. Paul, C. L. and Clark, S. J. Cytosine methylation: quantitation by automated genomic sequencing and GENESCAN analysis. *Biotechniques* **21:**126–133, 1996.
60. Herman, J. G., Graff, J. R., Myohanen, S., Nelkin, B. D., and Baylin, S. B. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc. Natl. Acad. Sci. USA* **93:**9821–9826, 1996.
61. Wu, D. Y., Ugozzoli, L., Pal, B. K., and Wallace, R. B. Allele-specific enzymatic amplification of beta-globin genomic DNA for diagnosis of sickle cell anemia. *Proc. Natl. Acad. Sci. USA* **86:**2757–2760, 1989.
62. Newton, C. R., Graham, A., Heptinstall, L. E., et al. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.* **17:**2503–2516, 1989.
63. Sarkar, G., Cassady, J., Bottema, C., and Sommer, S. S. Characterization of polymerase chain reaction amplification of specific alleles. *Anal. Biochem.* **186:**64–68, 1998.
64. Dodge, J. E., List, A. F., and Futscher, B. W. Selective variegated methylation of the p15 CpG island in acute myeloid leukemia. *Int. J. Cancer* **78:**561–567, 1998.
65. Tan, L. W., Bianco, T., and Dobrovic, A. Variable promoter region CpG island methylation of the putative tumor suppressor gene Connexin 26 in breast cancer. *Carcinogenesis* **23:**231–236, 2002.

66. Aggerholm, A. and Hokland, P. DAP-kinase CpG island methylation in acute myeloid leukemia: methodology versus biology? *Blood* **95:**2997–2999, 2000.
67. Aggerholm, A., Guldberg, P., Hokland, M., and Hokland, P. Extensive intra- and interindividual heterogeneity of p15INK4B methylation in acute myeloid leukemia. *Cancer Res*. **59:**436–441, 1999.
68. Xiong, Z. and Laird, P. W. COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res*. **25:**2532–2534, 1997.
69. Eads, C. A. and Laird, P. W. Combined bisulfite restriction analysis (COBRA). *Methods Mol. Biol*. **200:**71–85, 2002.
70. Sadri, R. and Hornsby, P. J. Rapid analysis of DNA methylation using new restriction enzyme sites created by bisulfite modification. *Nucleic Acids Res*. **24:**5058–5059, 1996.
71. Uejima, H., Lee, M. P., Cui, H., and Feinberg, A. P. Hot-stop PCR: a simple and general assay for linear quantitation of allele ratios. *Nat. Genet*. **25:**375–376, 2000.
72. Kuppuswamy, M. N., Hoffmann, J. W., Kasper, C. K., Spitzer, S. G., Groce, S. L., and Bajaj, S. P. Single nucleotide primer extension to detect genetic diseases: experimental application to hemophilia B (factor IX) and cystic fibrosis genes. *Proc. Natl. Acad. Sci. USA* **88:**1143–1147, 1991.
73. Gonzalgo, M. L. and Jones, P. A. Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). *Nucleic Acids Res*. **25:**2529–2531, 1997.
74. Gonzalgo, M. L. and Jones, P. A. Quantitative methylation analysis using methylation-sensitive single-nucleotide primer extension (Ms-SNuPE). *Methods* **27:**128–133, 2002.
75. El-Maarri, O., Herbiniaux, U., Walter, J., and Oldenburg, J. A rapid, quantitative, non-radioactive bisulfite-SNuPE- IP RP HPLC assay for methylation analysis at specific CpG sites. *Nucleic Acids Res*. **30:**e25, 2002.
76. Galm, O., Rountree, M. R., Bachman, K. E., Jair, K. W., Baylin, S. B., and Herman, J. G. Enzymatic regional methylation assay: a novel method to quantify regional CpG methylation density. *Genome Res*. **12:**153–157, 2002.
77. Cameron, E. E., Baylin, S. B., and Herman, J. G. p15(INK4B) CpG island methylation in primary acute leukemia is heterogeneous and suggests density as a critical factor for transcriptional silencing. *Blood* **94:**2445–2451, 1999.
78. Burri, N. and Chaubert, P. Complex methylation patterns analyzed by single-strand conformation polymorphism. *Biotechniques* **26:**232–234, 1999.
79. Bianco, T., Hussey, D., and Dobrovic, A. Methylation-sensitive, single-strand conformation analysis (MS-SSCA): A rapid method to screen for and analyze methylation. *Hum. Mutat*. **14:**289–293, 1999.
80. Maekawa, M., Sugano, K., Ushiama, M., et al. Heterogeneity of DNA methylation status analyzed by bisulfite–PCR–SSCP and correlation with clinico-pathological characteristics in colorectal cancer. *Clin. Chem. Lab. Med*. **39:**121–128, 2001.
81. Dobrovic, A., Bianco, T., Tan, L. W., Sanders, T., and Hussey, D. Screening for and analysis of methylation differences using methylation-sensitive single-strand conformation analysis. *Methods* **27:**134–138, 2002.
82. Liu, Q., Feng, J., Buzin, C., et al. Detection of virtually all mutations-SSCP (DOVAM-S): a rapid method for mutation scanning with virtually 100% sensitivity. *Biotechniques* **26:**932, 936–938, 940–942, 1999.
83. Guldberg, P., Worm, J., and Gronbaek, K. Profiling DNA methylation by melting analysis. *Methods* **27:**121–127,2002.
84. Worm, J., Aggerholm, A., and Guldberg, P. In-tube DNA methylation profiling by fluorescence melting curve analysis. *Clin. Chem*. **47:**1183–1189, 2001.
85. Baumer, A., Wiedemann, U., Hergersberg, M., and Schinzel, A. A novel MSP/DHPLC method for the investigation of the methylation status of imprinted genes enables the molecular detection of low cell mosaicisms. *Hum. Mutat*. **17:**423–430, 2001.
86. Deng, D. J., Zhou, J., Zhu, B. D., Ji, J. F., Harper, J. C., and Powell, S. M. Silencing-specific methylation and single nucleotide polymorphism of hMLH1 promoter in gastric carcinomas. *World J. Gastroenterol*. **9:**26–29, 2003.
87. Eads, C. A., Danenberg, K. D., Kawakami, K., et al. MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res*. **28:**E32, 2000.
88. Trinh, B. N., Long, T. I., and Laird, P.,W. DNA methylation analysis by MethyLight technology. *Methods* **25:**456–462, 2001.
89. Rand, K., Qu, W., Ho, T., Clark, S. J., and Molloy, P. Conversion-specific detection of DNA methylation using real-time polymerase chain reaction (ConLight-MSP) to avoid false positives. *Methods* **27:**114–120, 2002.
90. Rush, L. J. and Plass, C. Restriction landmark genomic scanning for DNA methylation in cancer: past, present and future applications. *Anal. Biochem*. **307:**191–201, 2002.
91. Costello, J. F., Plass, C., and Cavenee, W. K. Restriction landmark genome scanning. *Methods Mol. Biol*. **200:**53–70, (2002).
92. Huang, T. H., Perry, M. R. and Laux, D. E. Methylation profiling of CpG islands in human breast cancer cells. *Hum. Mol. Genet*. **8:**459–470, 1999.
93. Balog, R. P., Emi Ponce de Souza, Y., Tang, H. M., et al. Parallel assessment of CpG methylation by two-color hybridization with oligonucleotide arrays. *Anal. Biochem*. **309:**301–310, 2002.
94. Shi, H., Maier, S., Nimmrich, I., et al. Oligonucleotide-based microarray for DNA methylation analysis: principles and applications. *J. Cell Biochem*. **88:**138–143, 2003.
95. Li, L. C. and Dahiya, R. MethPrimer: designing primers for methylation PCRs. *Bioinformatics* **18:**1427–1431, 2002.
96. Singal, R. and Grimes, S. R. Microsoft Word macro for analysis of cytosine methylation by the bisulfite deamination reaction. *Biotechniques* **30:**116–120, 2001.
97. Esteller, M., Corn, P. G., Baylin, S. B., and Herman, J. G. A gene hypermethylation profile of human cancer. *Cancer Res*. **61:**3225–3229, 2001.
98. Laird, P. W. Early detection: the power and the promise of DNA methylation markers. *Nat. Rev. Cancer*. **3:**253–266, 2003.
99. Lee, W. H., Isaacs, W. B., Bova, G. S., and Nelson, W. G. CG island methylation changes near the GSTP1 gene in prostatic carcinoma cells detected using the polymerase chain reaction: a new prostate cancer biomarker. *Cancer Epidemiol. Biomarkers Prev*. **6:**443–450, 1997.
100. Wong, I. H., Lo, Y. M., and Johnson, P. J. Epigenetic tumor markers in plasma and serum: biology and applications to molecular diagnosis and disease monitoring. *Ann. N Y Acad. Sci*. **945:**36–50, 2001.
101. Jahr, S., Hentze, H., Englisch, S.,et al. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res*. **61:**1659–1665, 2001.
102. Belinsky, S. A., Nikula, K. J., Palmisano, W. A., et al. Aberrant methylation of p16(INK4a) is an early event in lung cancer and a potential biomarker for early diagnosis. *Proc. Natl. Acad. Sci. USA* **95:**11,891–11,896, 1998.
103. Ahrendt, S. A., Chow, J. T., Xu, L. H., et al. Molecular detection of tumor cells in bronchoalveolar lavage fluid from patients with early stage lung cancer. *J. Natl. Cancer Inst*. **91:**332–339, 1999.
104. Goessl, C., Muller, M., Straub, B., and Miller, K. DNA alterations in body fluids as molecular tumor markers for urological malignancies. *Eur. Urol*. **41:**668–676, 2002.
105. Evron, E., Dooley, W. C., Umbricht, C. B., et al. Detection of breast cancer cells in ductal lavage fluid by methylation-specific PCR. *Lancet* **357:**1335–1336, 2001.
106. Vogelstein, B., Fearon, E. R., Hamilton, S. R., et al. Clonal analysis using recombinant DNA probes from the X-chromosome. *Cancer Res*. **47:**4806–48013, 1987.
107. van Kamp, H., Jansen, R., Willemze, et al. Studies on clonality by PCR analysis of the PGK-1 gene. *Nucleic Acids Res*. **19:**2794, 1991.
108. Uchida, T., Ohashi, H., Aoki, E., et al. Clonality analysis by methylation-specific PCR for the human androgen-receptor gene (HUMARA-MSP). *Leukemia* **14:**207–212, 2000.
109. Nakahara, Y., Suzuki, H., Ohashi, H., et al. Clonality analysis of granulocytes and T lymphocytes in healthy females by the PCR-based HUMARA method. *Int. J. Hematol*. **69:**237-243.
110. Kubota, T., Das, S., Christian, S. L., Baylin, S. B., Herman, J. G. and Ledbetter, D. H. (1997) Methylation-specific PCR simplifies imprinting analysis. *Nat. Genet*. **16:**16–17, 1999.

111. Zeschnigk, M., Lich, C., Buiting, K., Doerfler, W., and Horsthemke, B. A single-tube PCR test for the diagnosis of Angelman and Prader–Willi syndrome based on allelic methylation differences at the SNRPN locus. *Eur. J. Hum. Genet*. **5:**94–98, 1997.
112. Kosaki, K., McGinniss, M. J., Veraksa, A. N., McGinnis, W. J., and Jones, K. L. Prader–Willi and Angelman syndromes: diagnosis with a bisulfite-treated methylation-specific PCR method. *Am. J. Med. Genet*. **73:**308–313, 1997.
113. Baumer, A. Analysis of the methylation status of imprinted genes based on methylation-specific polymerase chain reaction combined with denaturing high-performance liquid chromatography. *Methods* **27:**139–143, 2002.
114. Couvert, P., Poirier, K., Carrie, A., DHPLC-based method for DNA methylation analysis of differential methylated regions from imprinted genes. *Biotechniques* **34:**356–362, 2003.
115. Esteller, M., Hamilton, S. R., Burger, P. C., Baylin, S. B., and Herman, J. G. Inactivation of the DNA repair gene O^6-methylguanine-DNA methyltransferase by promoter hypermethylation is a common event in primary human neoplasia. *Cancer Res*. **59:**793–797, 1999.
116. Esteller, M., Garcia-Foncillas, J., Andion, E.,et al. Inactivation of the DNA-repair gene MGMT and the clinical response of gliomas to alkylating agents. *N. Engl. J. Med*. **343:**1350–1354, 2000.
117. Esteller, M., Gaidano, G., Goodman, S. N., et al. Hypermethylation of the DNA repair gene O(6)-methylguanine DNA methyltransferase and survival of patients with diffuse large B-cell lymphoma. *J. Natl. Cancer Inst*. **94:**26–32, 2002.
118. Li, G. M. The role of mismatch repair in DNA damage-induced apoptosis. *Oncol. Res.* **11:**393–400, 1999.
119. Elsaleh, H., Shannon, B., and Iacopetta, B. Microsatellite instability as a molecular marker for very good survival in colorectal cancer patients receiving adjuvant chemotherapy. *Gastroenterology* **120:**1309–1310, 2001.
120. Hemminki, A., Mecklin, J. P., Jarvinen, H., Aaltonen, L. A., and Joensuu, H. Microsatellite instability is a favorable prognostic indicator in patients with colorectal cancer receiving chemotherapy. *Gastroenterology* **119:**921–928, 2000.
121. Whitehall, V. L., Walsh, M. D., Young, J., Leggett, B. A., and Jass, J. R. Methylation of O-6-methylguanine DNA methyltransferase characterizes a subset of colorectal cancer with low-level DNA microsatellite instability. *Cancer Res*. **61:**827–830, 2001.
122. Jass, J. R., Walsh, M. D., Barker, M., Simms, L. A., Young, J., and Leggett, B.A. Distinction between familial and sporadic forms of colorectal cancer showing DNA microsatellite instability. *Eur. J. Cancer* **38:**858–866, 2002.
123. Boland, C. R., Thibodeau, S. N., Hamilton, S. R., et al. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res*. **58:**5248–5257, 1998.
124. Warnecke, P. M., Stirzaker, C., Melki, J. R., Millar, D. S., Paul, C. L., and Clark, S. J. Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. *Nucleic Acids Res*. **25:** 4422–4426, 1997.

IV OTHER CLINICAL DIAGNOSTIC TECHNOLOGIES

13 Flow Cytometry

JOSEPH A. DIGIUSEPPE

1. GENERAL PRINCIPLES OF FLOW CYTOMETRY

Flow cytometry is a technique in which single cells in a fluid suspension are analyzed with respect to their intrinsic light-scattering properties and are simultaneously evaluated for one or more extrinsic properties (i.e., the presence of specific molecules) using fluorescent probes. The fluorescent probe might bind directly to the targeted molecule (e.g., propidium iodide in DNA content analysis) or a fluorescent dye might be coupled to an antibody probe, to enable detection of a specific protein. By using several different fluorochromes with substantially nonoverlapping emission spectra, the laboratorian can simultaneously evaluate the expression of multiple extrinsic cellular properties. Moreover, because the flow rate of cells within the flow cytometer is rapid, thousands of cells can be analyzed in seconds. Despite the rapidity with which data can be acquired, however, because each cell is analyzed individually, multiple intrinsic and extrinsic parameters are retained for each cell; that is, flow cytometric analysis is multiparametric. In this section, we review some general principles of flow cytometry *(1–6)*.

1.1. FLUIDICS A requirement of flow cytometry is that the sample to be analyzed must comprise single cells in a fluid suspension. Therefore, hematopoietic precursors and lymphocytes are ideally suited for flow cytometric analysis, because they normally exist as single-cells (i.e., without intercellular attachments) in vivo. The single-cell suspension is aspirated into the flow cytometer, whereupon it encounters an isotonic fluid referred to as sheath fluid. The sheath fluid surrounds the sample fluid and produces conditions of laminar flow within the stream of sample fluid. This process of hydrodynamic focusing results in a single-file arrangement of cells within the sample stream.

1.2. LIGHT SOURCE AND LIGHT SCATTER Next, each cell is interrogated by a light source. Typically, clinical flow cytometers use small air-cooled lasers as a light source. The wavelength of monochromatic light emitted by the laser (e.g., 488 nm for argon) in turn dictates the fluorochromes suitable for use with a given instrument; that is, a useful fluorochrome must absorb significantly at that wavelength. As each cell passes through the interrogation point, it scatters the incident laser light in all directions. Light scattered at two specific angles is measured by the flow cytometer: forward-angle light scatter (FSC), and orthogonal or right-angle light scatter (SSC). FSC refers to light deflected by the cell at a small angle relative to the vector of the incident light, whereas SSC refers to light deflected at a right angle to the vector of the incident light. Appropriately arrayed photodetectors capture the light scattered by the cell at low (FSC) or right (SSC) angles and convert its energy into an electrical signal proportional to the light's intensity. This electrical signal is ultimately converted into a digital signal, which is stored by a dedicated computer system. To a rough approximation, the FSC signal of a cell correlates with its size, whereas the SSC signal correlates with granularity or other intracellular "complexity" (e.g., vacuoles) (Fig. 1).

1.3. FLUORESCENCE AND EXTRINSIC PROPERTIES Unlike FSC and SSC, which represent light-scattering properties intrinsic to the cell, extrinsic properties require the addition of a fluorescent probe for their measurement. A fluorescent molecule is one that absorbs light across a spectrum of wavelengths and emits light of lower energy across a spectrum of longer wavelengths. The difference between the peak absorption and emission wavelengths of a fluorochrome is referred to as its Stokes' shift. By using a series of fluorochromes with progressively larger Stokes' shifts, each of which absorbs reasonably well at the wavelength of the laser used in the flow cytometer, the laboratorian can simultaneously evaluate the cell for several extrinsic properties. The clinical utility of such multicolor analysis is enhanced when the fluorescent data are analyzed in conjunction with FSC and SSC.

1.4. MULTICOLOR ANALYIS The number of fluorochromes capable of being used simultaneously is limited by the number of photodetectors in the flow cytometer. The specificity of each photodetector for a given band of wavelengths results from the arrangement of a series of mirrors and filters, which permits only light of certain wavelengths to reach each detector. Most clinical flow cytometers in use today are capable of three-color or four-color analysis. However, the range of suitable fluorochromes can be expanded by adding additional light sources whose emission differs from that of the primary laser and by using tandem dyes. The latter consist of two coupled fluorescent molecules; the first is excited by the light source

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

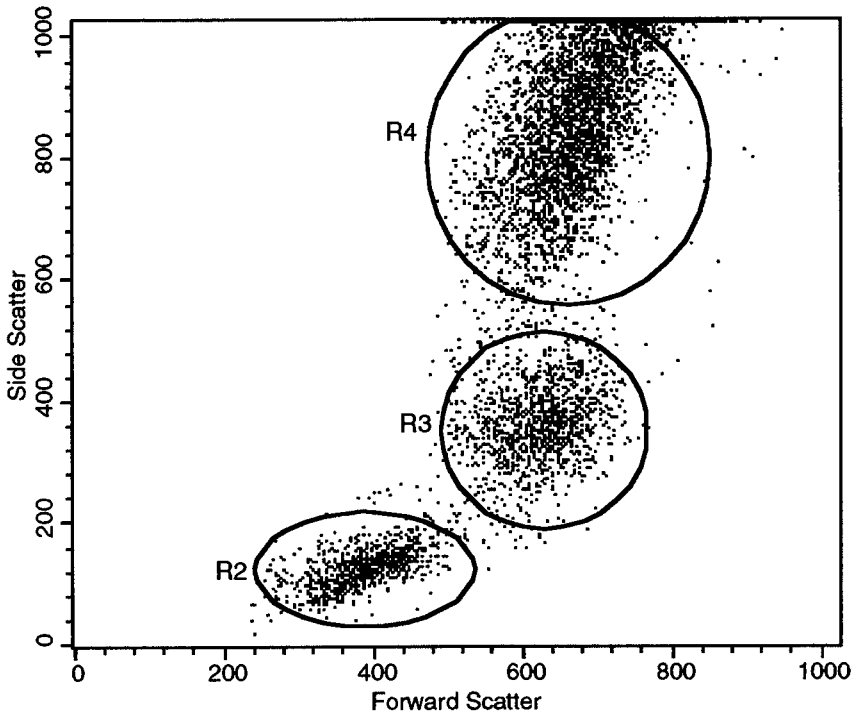


Fig. 1. Normal peripheral blood specimen illustrating the intrinsic light-scatter properties of lymphocytes (R2), monocytes (R3), and granulocytes (R4).

and the second is excited by the emission of the first. In this way, a larger net Stokes' shift is produced compared with individual fluorochromes, and more extrinsic properties can be measured using a single light source. By using multiple light sources and judicious combinations of fluorochromes, including tandem dyes, investigators have developed research instruments capable of 11-color analysis *(7)*.

1.5. SPECTRAL OVERLAP AND COMPENSATION One pitfall of multicolor analysis is spectral overlap. Because the emission spectra of individual fluorochromes are often broad and overlapping, a portion of the fluorescence signal arriving at a specific photodetector could have originated from a different fluorochrome. The contribution of such extraneous signals can be measured by omitting individual fluorochromes in multicolor experiments and subtracting the fraction of signal originating from inappropriate fluorochromes. For example, the "compensated" signal for photodetector number 2 might be expressed as the total signal recorded in photodetector number 2, minus 20% of the signal recorded in photodetector number 1 (i.e., compensated FL2 = [FL2] – 0.2[FL1]), whereas the "compensated" signal for photodetector number 1 might be expressed as the total signal recorded in photodetector number 1, minus 5% of the signal recorded in photodetector number 2 (i.e., compensated FL1 = [FL1] – 0.05[FL2]) (Fig. 2). Compensation is commonly performed prior to acquisition of the actual experimental data, but software has more recently become available that enables postacquisition compensation of raw (uncompensated) data.

1.6. DATA ANALYSIS AND GATING Once the intrinsic and extrinsic cellular properties of many cells (typically 5000–20,000 in routine clinical studies) have been measured and recorded in the form of a list (so-called "list-mode" data), the laboratorian must analyze the data. Such multiparametric analysis invariably involves electronic gating. When primary data are displayed (typically as one-dimensional histograms, or two-dimensional dot plots, with each dot, or event, representing a single cell), specific populations can be selected for analysis to the exclusion of other (irrelevant) populations. For example, a polygonal or amorphous gate might be drawn around events positive for the B-cell antigen CD20 in the case of a B-cell non-Hodgkin's lymphoma (Fig. 3). Gating enables precise characterization of the extrinsic properties of specific subsets of cells, without contamination from fluorescence signals originating from cells not relevant to a particular analysis.

2. CLINICAL APPLICATIONS OF FLOW CYTOMETRY

Routine applications of flow cytometry in the clinical laboratory include enumeration of peripheral blood CD4+ T-cells in patients with human immunodeficiency virus (HIV) infection, enumeration of CD34+ stem cells in peripheral blood and bone marrow for use in stem cell transplantation, and immunophenotypic characterization of acute leukemias, non-Hodgkin's lymphomas, and other lymphoproliferative disorders. In certain instances, the DNA content, or ploidy status of tumor cells, which is readily measured by flow cytometry using a DNA-binding dye, might have prognostic value. In recent years, flow cytometry has also been used to monitor patients with leukemia or lymphoma who have already been treated and have attained a clinical remission. The detection of so-called "minimal residual disease" by flow cytometry in these patients predicts a higher likelihood of relapse and might permit more precise tailoring of additional therapy.

2.1. CD4+ T-CELL ENUMERATION AND OTHER ASSAYS OF IMMUNE DEFICIENCY A principal mechanism whereby HIV abolishes normal immune function is infection and depletion of a subset of T-cells bearing the CD4

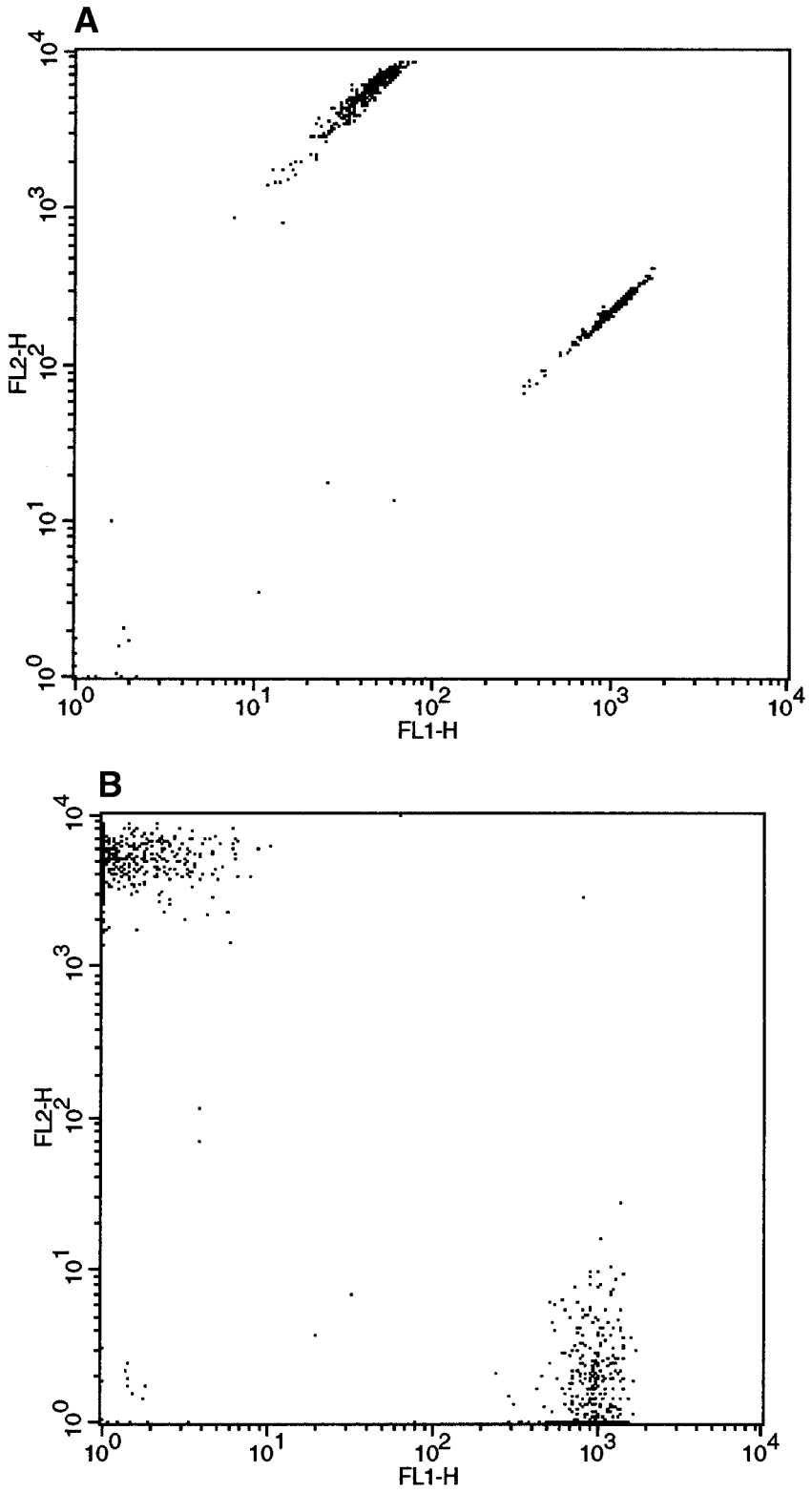


Fig. 2. Fluorescence signals obtained using two fluorochromes (FITC, FL1; PE, FL2) with overlapping emission spectra before **(A)** and after **(B)** compensation.

molecule *(8–10)*. The extent of depletion of CD4+ T-cells in HIV-infected patients correlates with prognosis. Moreover, in conjunction with HIV viral load measurement, CD4+ T-cell enumeration provides a quantitative assay for the efficacy of therapy in patients treated with antiretrovirals. Traditionally, flow cytometry has been used to calculate the percentage of white blood cells that are T-cells (CD3+) that also express the T-cell subset antigen CD4 (Fig. 4). In this methodology, the absolute number of CD4+ T-cells is calculated by multiplying the absolute lymphocyte count (measured using a routine

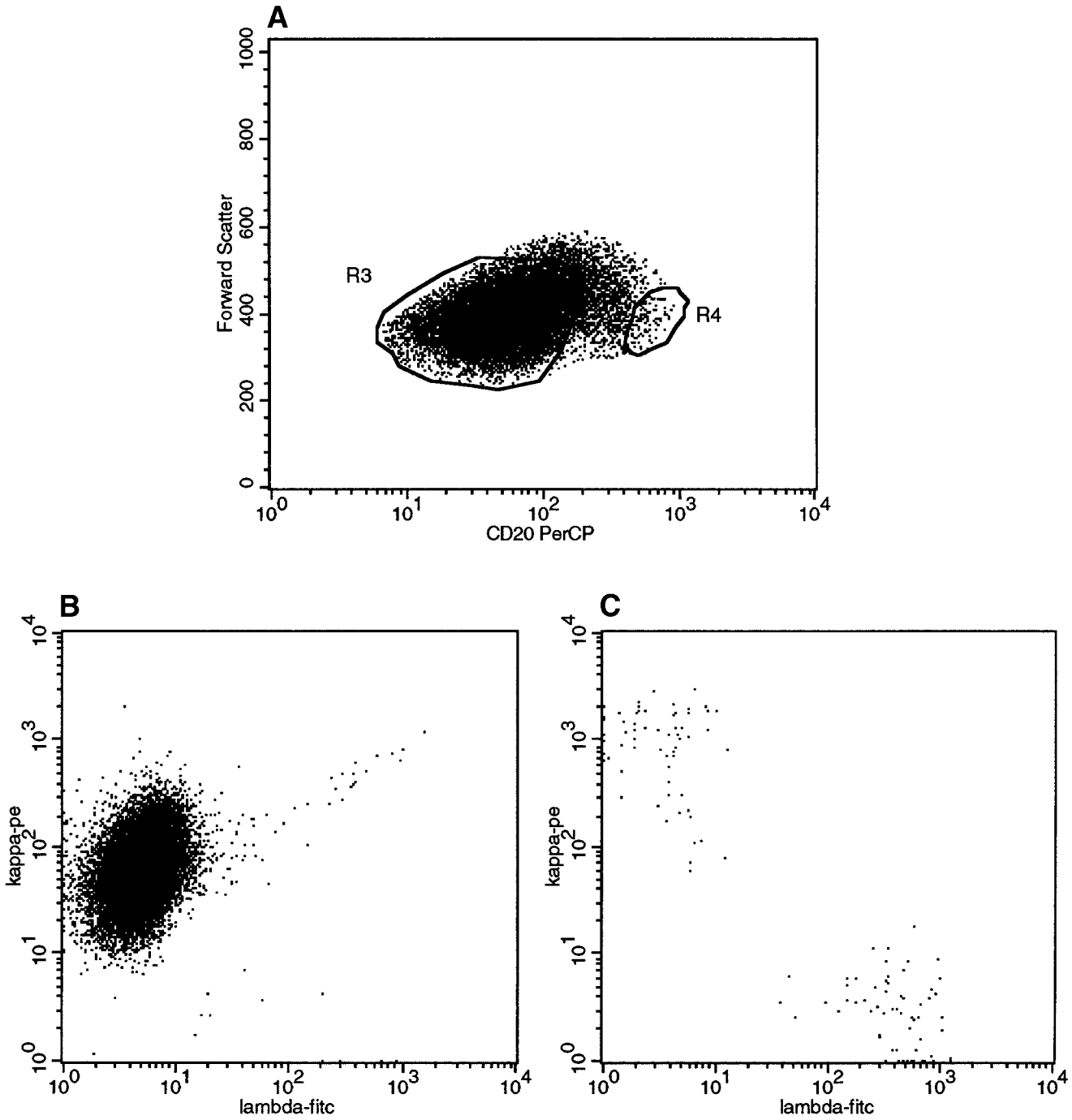


Fig. 3. Example of electronic gating. **(A)** All B-cells are displayed and all are positive for the B-cell antigen CD20. Selective gating on B-cells whose CD20 expression is comparatively dim (R3) or bright (R4), however, discloses that the predominant population of abnormal CD20(dim)+ B-cells expresses monotypic kappa light chain **(B)** whereas the few remaining normal CD20(bright)+ B-cells comprise a mixture of kappa- and lambda-bearing cells (**C**).

hematology analyzer) by the percentage of CD4+ T-cells determined flow cytometrically.

Interlaboratory reproducibility of CD4+ T-cell enumeration can be enhanced by eliminating the requirement for separate measurement of the white blood cell count. In so-called single-platform methods, a precise volume of blood is added to a commercially prepared tube containing a known number of microfluorospheres as well as an appropriate combination of fluorochrome-conjugated antibodies. Because the total number of microfluorospheres is known, the absolute number of lymphocytes (which are characteristically brightly positive for CD45) can be calculated by comparison with the number of microfluorospheres detected during the acquisition period. This value is then multiplied by the percentage of CD4+ T-cells to yield the absolute CD4+ T-cell count.

It should be noted that flow cytometric evaluation of immune deficiency is not limited to acquired immunodeficiency. There are many forms of primary (congenital) immunodeficiency,

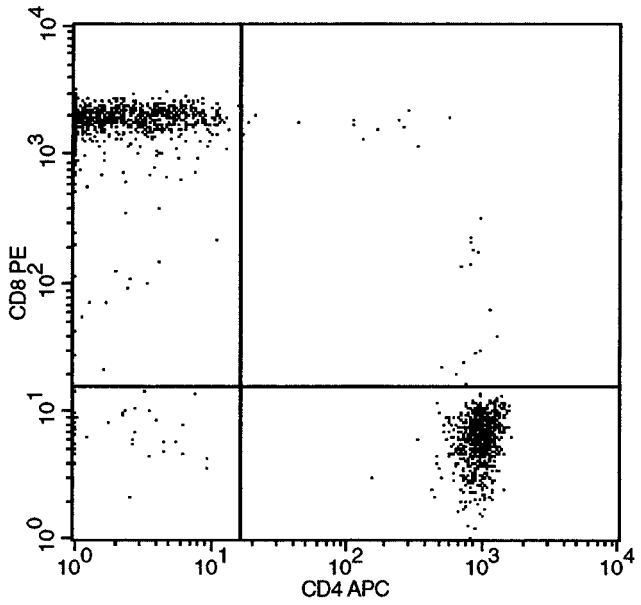


Fig. 4. Peripheral blood specimen, gated on CD3+ T-cells. In this manner, the percentage of CD4+ T-cells (lower right quadrant) and CD8+ T-cells (upper left quadrant) can be quantified.

including disorders of cellular immunity *(11)*. Lymphocyte subset analysis by flow cytometry is typically one of the first laboratory tests performed in the clinical evaluation of such patients. Moreover, lymphocyte activation, cytokine expression within specific cell subsets, and cellular cytotoxicity can all be probed using flow cytometry, although such functional assays are generally performed only in select laboratories with experience in these techniques.

2.2. CD34+ HEMATOPOIETIC STEM CELL ENUMERATION Hematopoietic stem cells are capable of reconstituting hematopoiesis; through multiple generations of proliferation and differentiation, a single pluripotent stem cell can give rise to red blood cells, white blood cells, and platelets. In patients with malignancies whose successful treatment requires extremely high doses of chemotherapy, a side effect of therapy is ablation of the bone marrow (so-called myeloablative therapy), the normal site of hematopoiesis. Without reconstitution of hematopoiesis by replacement of stem cells, either harvested from the patient prior to therapy (autologous) or from an immunologically compatible donor (allogeneic), such a potentially curative therapy would be fatal. Allogeneic stem cell transplant might also be used to cure patients with non-neoplastic, but potentially fatal genetic disorders (e.g., severe combined immunodeficiency). In either case, successful stem cell engraftment requires delivery of an adequate "dose" of stem cells to the recipient.

The expression of CD34 by hematopoietic stem cells, which can be obtained from bone marrow, peripheral blood, or umbilical cord blood samples, enables their rapid enumeration by flow cytometry *(12,13)*. As in CD4+ T-cell enumeration, both dual-platform (requiring both flow cytometry and a conventional hematology analyzer) and single-platform (bead-based) assays are in use clinically. Several general points can be made, however, with respect to CD34+ stem cell enumeration. Because CD34+ stem cells are generally rare (<1%), a large number of total events must be collected to enable their precise quantification. In addition, antibodies to CD34 whose binding requires recognition of carbohydrates and terminal sialic acid molecules (so-called class I antibodies) are to be avoided, as these post-translational modifications of CD34 are only variably present. Finally, it is preferable to use an antibody conjugated to the fluorochrome phycoerythrin (PE), as its relatively high quantum yield, or brightness, enhances the sensitivity of the assay for stem cells, that are only dimly positive for CD34.

2.3. IMMUNOPHENOTYPIC ANALYSIS OF ACUTE LEUKEMIA The acute leukemias represent malignancies of lymphoid (acute lymphoblastic leukemia [ALL]) or myeloid (acute myeloid leukemia [AML]) precursors *(14)*. Patients with acute leukemia commonly present with signs and symptoms of bone marrow failure (e.g., fatigue, bruising, anemia, thrombocytopenia) and, not infrequently, with an elevated white blood cell count. However, despite the many clinical similarities between ALL and AML, optimal treatment differs for these two broad categories of acute leukemia. Moreover, both ALL and AML encompass clinically and biologically distinct subsets of disease, which are often impossible to discern on the basis solely of conventional microscopic analysis. Because immunologic characterization of the neoplastic cells (blasts) in many cases permits more precise, therapeutically relevant diagnostic information about an acute leukemia than conventional microscopic evaluation, flow cytometry has become an important adjunct in the diagnosis of patients with acute leukemia.

2.3.1. Acute Myeloid Leukemia In AML, the neoplastic cells, or blasts, are myeloid precursor cells that might display varying degrees of differentiation to granulocytes, monocytes, or, less commonly, erythroid or megakaryocytic precursors. It is important to remember though that diagnostic specimens (typically peripheral blood or bone marrow) contain mature myeloid cells in addition to the malignant blasts. Because myeloblasts can share many of the immunophenotypic properties of these mature myeloid cells, it is important to gate selectively upon the blasts. Historically, blasts were gated on the basis only of their FSC and SSC signals. However, this approach is significantly flawed, because gates drawn solely on the basis of these intrinsic properties invariably include non-neoplastic lymphoid cells and erythroid precursors, as well as monocytes; the "population" gated using only FSC and SSC, therefore, in fact comprises a heterogeneous mixture of non-neoplastic and malignant cells. In the last decade, it has become routine to gate on blasts using the combination of dim CD45 expression and low SSC *(15,16)*. This approach permits more precise distinction of blasts from other bone marrow elements (Fig. 5). In cases of acute leukemia, blasts gated in this fashion are substantially free of contaminating non-neoplastic cells (Fig. 6).

Once a candidate blast population has been isolated by electronic gating, it can be probed immunophenotypically using a panel of antibodies *(15–17)*. It is important to include antibodies recognizing not only myeloid antigens but also B-lymphoid and T-lymphoid antigens, as so-called "aberrant" expression of myeloid antigens in ALL and lymphoid antigens in AML is not uncommon. In most cases, it is possible to assign unambiguously a lineage (i.e., myeloid vs T- or B-lymphoid) on the basis

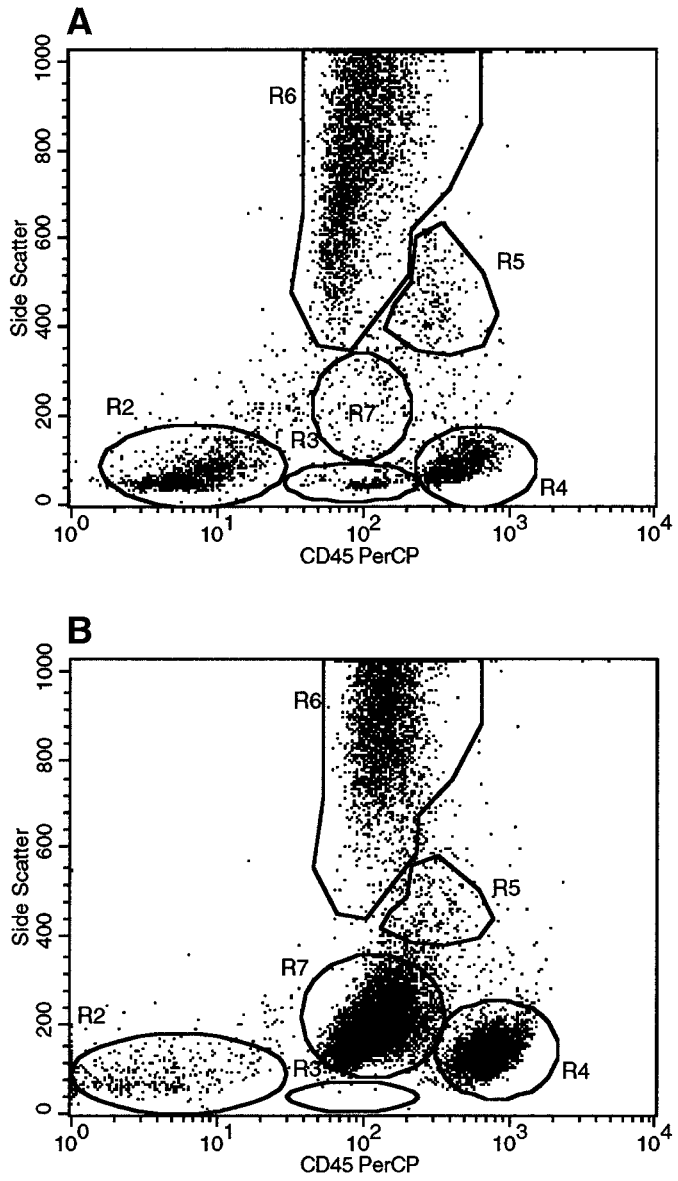


Fig. 5. CD45 expression vs SSC. Normal bone marrow (**A**) and bone marrow of a patient with acute leukemia (**B**). R2 = erythroid precursors; R3 = normal B-cell precursors (hematogones); R4 = lymphocytes; R5 = monocytes; R6 = granulocytes; R7 = blasts.

of the surface-membrane immunophenotype of the blasts. However, in some cases, it might be useful also to evaluate cytoplasmic antigens whose expression is more specific for a given lineage. For example, detection of cytoplasmic myeloperoxidase in a blast population would provide strong evidence for myeloid differentiation, even if that population were found to express lymphoid antigens on the cell membrane. Some of the antigens commonly evaluated to permit lineage assignment in cases of acute leukemia are summarized in Table 1.

In addition to facilitating lineage assignment, flow cytometric immunophenotyping in AML can also provide more detailed information with respect to the type (e.g., granulocytic, monocytic, megakaryocytic, erythroid) and extent of myeloid differentiation. Perhaps more importantly, certain composite immunophenotypes have been associated with specific recurrent karyotypic abnormalities in AML, some of which require modification of the patient's treatment protocol. For example, in acute promyelocytic leukemia, absent or weak expression of HLA-DR is typical. This form of AML must be specifically recognized, because its characteristic juxtaposition of PML and RARα genes resulting from a translocation between chromosomes 15 and 17 renders the leukemic cells susceptible to all-*trans* retinoic acid. Another biologically distinct subset of AML, characterized karyotypically by the t(8;21) AML1-ETO is associated with a CD34+, HLA-DR+, myeloid immunophenotype, with aberrant expression of the B-cell antigen CD19 and at least partial positivity for the nuclear antigen terminal transferase (TdT). Finally, AML with abnormalities of 16q, most commonly inv(16), is associated with aberrant expression of the T-cell antigen CD2.

2.3.2. Acute Lymphoblastic Leukemia The term ALL includes precursor B-cell and precursor T-cell forms of the disease. Despite differences in clinical presentation and optimal treatment, precursor B-cell ALL and precursor T-cell ALL are morphologically indistinguishable *(18)*. Flow cytometric immunophenotyping is, therefore, essential in establishing the diagnosis of ALL. As with AML, a panel of antibodies should be used, including B-lineage, T-lineage, and myeloid antigens *(16,18–20)*. As noted earlier, surface-membrane myeloid antigens can be expressed aberrantly in ALL, a feature that might complicate lineage assignment of the blast population. In such instances, it is appropriate to evaluate the expression of cytoplasmic antigens more specifically associated with B- and T-lymphoid differentiation (i.e., cytoplasmic CD22 and cytoplasmic CD3, respectively).

In addition to its role in establishing the diagnosis of precursor B-cell ALL and precursor T-cell ALL, flow cytometric immunophenotyping might provide prognostic information and predict recurrent genotypic aberrations. For example, in childhood precursor B-cell ALL, bright expression of CD20 and CD45 are adverse prognostic factors, independent of other known risk factors, whereas in childhood precursor T-cell ALL, expression of the T-cell antigen CD2 is a favorable prognostic factor. A number of recurrent balanced translocations have also been associated with specific immunophenotypic profiles. Such composite immunophenotypes have been described in precursor B-cell ALL with t(1;19) E2A-PBX, t(12;21) TEL-AML1, and t(4;11) AF4-MLL.

3. IMMUNOPHENOTYPIC ANALYSIS OF NON-HODGKIN'S LYMPHOMA AND OTHER LYMPHOPROLIFERATIVE DISORDERS

The non-Hodgkin's lymphomas (NHLs) are malignancies of B-cells, T-cells, or, rarely, natural killer cells, which present clinically with swelling of lymph nodes and/or tumoral masses in extranodal sites *(14)*. In cases of NHL, the diagnostic specimen is often a biopsy of an affected lymph node or extranodal site. To make such a specimen suitable for flow cytometric immunophenotyping, the biopsy must be disaggregated and a single-cell suspension prepared; this is typically accomplished mechanically, although enzymatic digestion has also been used. Increasingly, tumor cells from both superficial and deep-seated lesions (the latter is performed with radiographic guidance) can

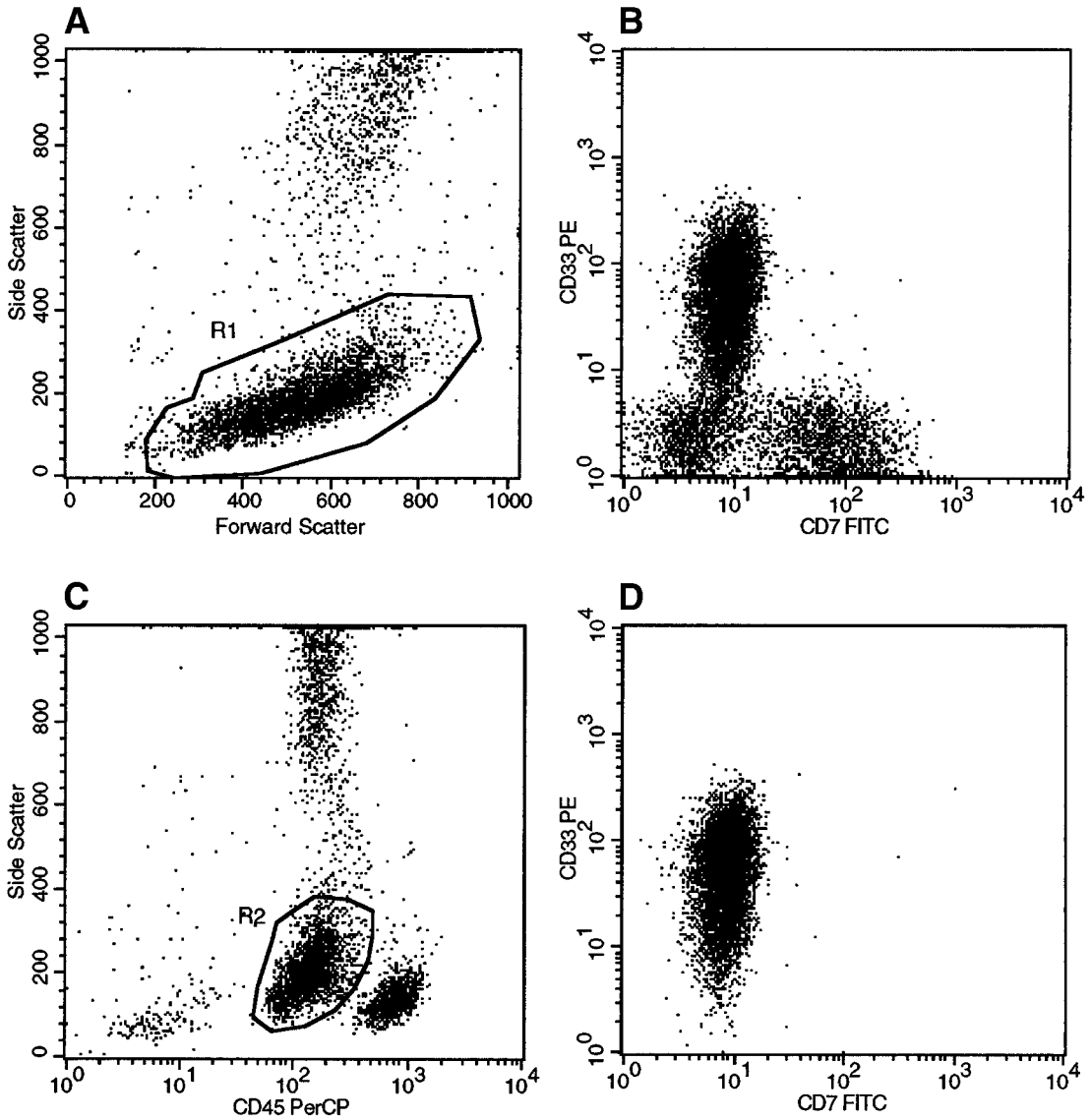


Fig. 6. Illustration of CD45 vs SSC gating in acute leukemia. In **(A)**, a conventional mononuclear cell gate is drawn (R1) on the basis of intrinsic cellular light-scatter properties. The fluorescence signals obtained using this strategy (shown in **B**) are heterogeneous, as the gate includes not only blasts but also non-neoplastic erythroid precursors, T-cells, B-cells, and monocytes. Because blasts characteristically express CD45 dimly (R2 **[C]**), the fluorescence signals obtained using this approach **(D)** are essentially homogeneous. In this case, the blasts are positive for the myeloid antigen CD33 **(B, D)**, whereas the non-neoplastic cells contaminating the R1 gate are negative for CD33 **(B)**.

Table 1
Some Lineage-Related Antigens Commonly Evaluated in Acute Leukemia

Myeloid	*T-lymphoid*	*B-lymphoid*
CD11b	CD1a	CD19
CD13	CD2	CD20
CD14	CD3	CD22
CD15	CD5	cytoCD22
CD33	CD7	
Myeloperoxidase	cytoCD3	

be aspirated directly using a narrow-gauge needle. Such fine-needle aspiration procedures not only spare the patient potentially unnecessary surgery, but also yield single-cell specimens suitable for flow cytometric immunophenotyping with a minimum of further preparation. Flow cytometry is also frequently used in the diagnosis of other chronic lymphoproliferative disorders, which are neoplasms of mature lymphocytes and typically present in the peripheral blood and/or bone marrow, rather than lymph nodes *(21–24)*.

3.1. B-CELL NON-HODGKIN'S LYMPHOMAS/LYMPHOPROLIFERATIVE DISORDERS B-cell NHLs comprise the majority of NHLs in Western countries. Flow cytometric immunophenotyping is a valuable diagnostic adjunct, both in distinguishing B-cell NHLs from benign lymphoid hyperplasia and in subclassification of B-cell NHLs. Mature B-cells display an antigen receptor (antibody) that consists of heavy chains and light chains (kappa or lambda). Because the antigen receptor expressed by an individual B-cell contains either a kappa or lambda light chain, a neoplastic B-cell population representing a monoclonal B-cell proliferation should contain exclusively

kappa- or lambda-bearing cells (e.g., Fig. 3B). In practice, a ratio of kappa-bearing : lambda-bearing B-cells in excess of 10 : 1 or less than 1 : 5 reflects the presence of a monoclonal B-cell population. For comparison, in (polyclonal) B-cell hyperplasia, the kappa : lambda ratio typically ranges between 1 : 1 and 3 : 1.

Not only does flow cytometric immunophenotyping facilitate the distinction between benign, reactive lymphoid proliferations, and monoclonal neoplasms, the extensive immunophenotypic profiles made possible by flow cytometry enable more precise distinction among different B-cell NHLs. For example, B-cell chronic lymphocytic leukemia and mantle cell lymphoma could be confused morphologically, but their predicted clinical courses differ significantly. Although both are typically neoplasms of CD5+ B-cells, there are characteristic differences in the presence and/or intensity of expression of several molecules, including CD20, CD23, FMC-7, and immunoglobulin light chain, which facilitate distinction in most cases. Hairy cell leukemia is another B-cell lymphoproliferative disorder whose diagnosis is simplified by flow cytometry. Although neoplastic cells might not be numerous in the peripheral blood, in the appropriate clinical setting the detection of a monoclonal B-cell population with the composite immunophenotype CD11c(bright)+, CD25+, CD103+ is virtually pathognomonic of hairy cell leukemia, which requires specific therapy.

3.2. T/NK-CELL NON-HODGKIN'S LYMPHOMAS/LYMPHOPROLIFERATIVE DISORDERS Whereas immunoglobulin light-chain restriction permits flow cytometric demonstration of B-cell monoclonality, there is no analagous immunophenotypic marker of clonality in T-cell or natural killer (NK)-cell NHLs, or lymphoproliferative disorders. Nonetheless, flow cytometry is useful in detecting immunophenotypic aberrations that are common in T-cell neoplasia. For instance, among the so-called peripheral (i.e., nonlymphoblastic) T-cell lymphomas, absent or diminished expression of CD7, a ubiquitous T-cell antigen, is extremely common. Moreover, in conjunction with the clinical and morphologic findings, a detailed immunophenotypic profile as determined by flow cytometry facilitates distinction among different T-cell or NK-cell NHLs/lymphoproliferative disorders.

4. DNA CONTENT/PLOIDY ANALYSIS

Propidium iodide (PI) is a DNA-intercalating dye that binds DNA stoichiometrically; the fluorescence intensity of PI, therefore, correlates with the amount of DNA contained within the nucleus. When a population of nuclei is stained with PI or another similar DNA-binding fluorescent dye, a histogram of DNA content reflects the proportions of cells in different phases of the cell cycle. In a number of hematopoietic and solid tumors, the fraction of tumor cells in the S-phase and/or the presence of an aneuploid cell population as determined by flow cytometry has been shown to be prognostic *(25)*. It is important to recognize that not all studies have demonstrated prognostic significance for S-phase fraction or DNA content measurement. In a few specific instances though, DNA ploidy measurements are still routinely performed for diagnostic or prognostic purposes. In gestational trophoblastic disease, for example, DNA content is commonly analyzed by flow cytometry (Fig. 7). In these proliferative disorders of placental trophoblast, a triploid DNA content correlates pathologically with a partial hydatidiform mole, which has no malignant potential, whereas a diploid DNA content correlates with a complete hydatidiform mole, which is associated with a markedly elevated risk for the subsequent development of choriocarcinoma, a malignancy of trophoblastic tissue *(26)*. DNA ploidy is also typically evaluated at the time of diagnosis in patients with precursor B-cell ALL, in which a hyperdiploid karyotype is associated with a favorable prognosis *(27)*.

5. MINIMAL RESIDUAL DISEASE ANALYSIS

Historically, it has been apparent that a significant portion of patients treated for acute leukemia who attain a complete clinical remission will, nonetheless, subsequently relapse. This observation implies the persistence of small (subclinical) amounts of disease following therapy in some patients, who are thus at elevated risk for recurrence. Patients with acute leukemia who have fewer than 5% blasts in the bone marrow following therapy are considered in morphologic remission. Although precise enumeration of blasts below this level is difficult by conventional microscopic evaluation, flow cytometric immunophenotyping is capable of detecting and precisely enumerating as few as 1 blast in 10,000 cells (i.e., 0.01% or 10^{-4}).

Because leukemic blasts commonly display one or more "aberrant" antigens or express appropriate antigens at densities different from those seen in normal hematopoietic precursors, it is possible immunophenotypically to distinguish a minor population of leukemic blasts from normal regenerating hematopoietic precursors *(28–30)*. In ALL, greater than 90% of cases are amenable to such immunophenotypic analysis, whereas in AML, approx 70–75% of cases can be evaluated immunophenotypically. In order to obtain sensitivity on the order of 10^{-4}, it is necessary to analyze a large total number of cells (e.g., 500,000), a practice that is feasible using flow cytometry. Advantages of flow cytometric compared with molecular detection of minimal residual disease are rapid turnaround time and ability to quantify minimal residual disease directly. In ALL, for example, molecular detection of minimal residual disease requires sequencing of clone-specific rearrangements of immunoglobulin or T-cell receptor genes. Moreover, polymerase chain reaction-based methodologies are necessarily only semiquantitative, and several studies have shown that the percentage of residual disease, not merely its presence or absence, is significant in predicting outcome. In both AML and pediatric ALL, several groups have shown that the detection of residual leukemic blasts immediately or at different times following remission induction is associated with a higher likelihood of subsequent relapse. These results hold the potential both to identify earlier patients who will require additional therapy and to identify particularly low-risk patients who might safely be spared the toxicity of additional therapy.

6. PAROXYSMAL NOCTURNAL HEMOGLOBINURIA

Paroxysmal nocturnal hemoglobinuria (PNH) is a rare, acquired clonal disorder of hematopoiesis associated clinically with hemolysis, thrombosis, and bone marrow failure

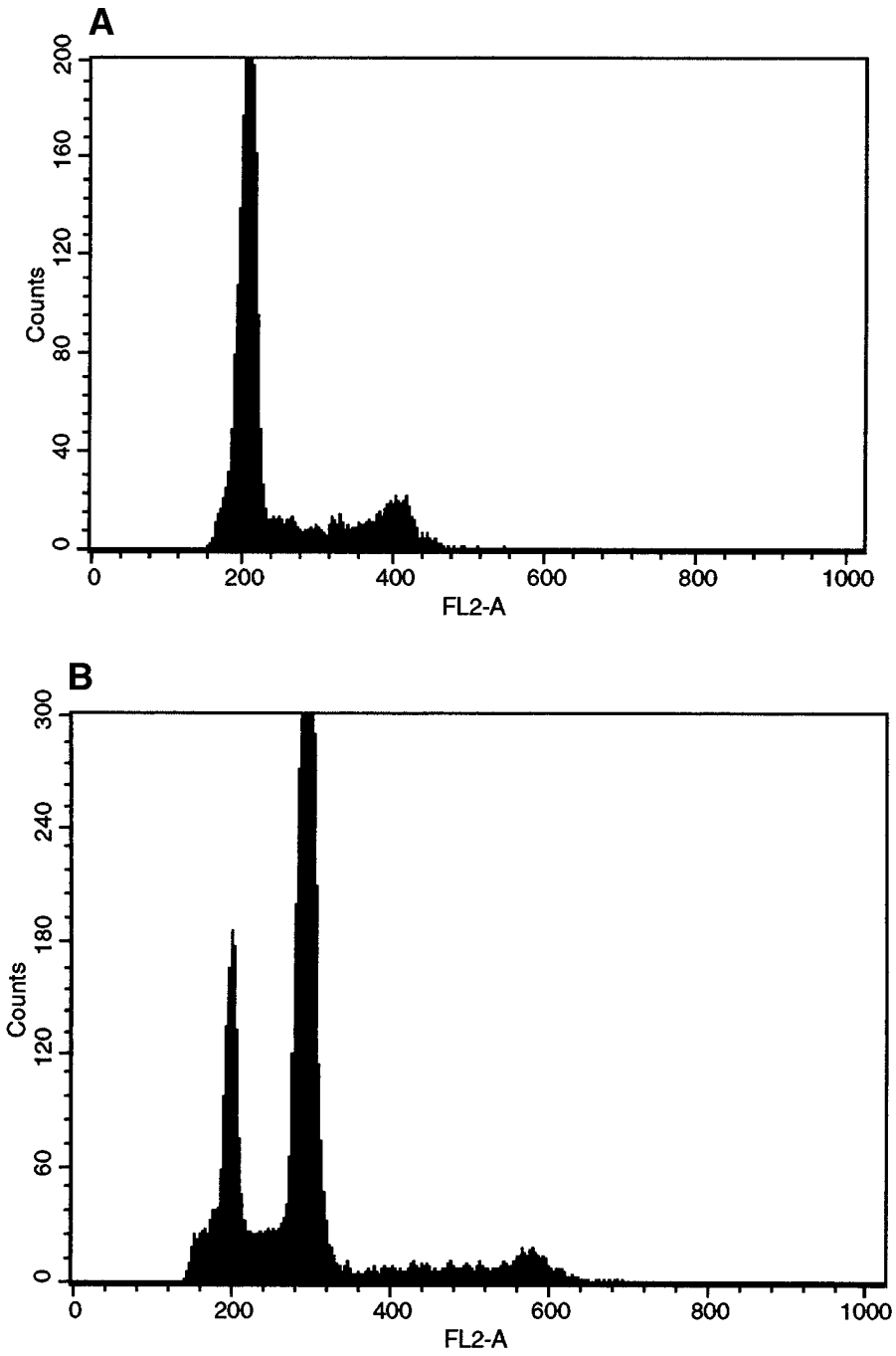


Fig. 7. DNA content as measured by binding of propidium iodide. **(A)** Histogram of normal diploid pattern. In this display, the modal fluorescence of cells with 2n DNA (i.e., G_1-phase) is about channel 200, whereas that of cells with 4n DNA (i.e., G_2-phase) is about channel 400. Cells engaged in DNA synthesis (S-phase) are found between these two peaks. **(B)** Partial hydatidiform mole. The G_1 peak of normal cells is seen at about channel 200, whereas the G_1 peak of the aneuploid (triploid, or 3n) cell population is seen at about channel 300. The peak at about channel 600 represents the G_2 peak of the aneuploid population (6n DNA).

syndromes (aplastic anemia and myelodysplasia). This disorder results from an acquired mutation in PIG-A, whose gene product is necessary for the synthesis of glycosylphosphatidyl inositol (GPI), a molecule required for retention of many membrane-bound proteins. Because some of the membrane proteins critical in regulating erythrocyte susceptibility to complement-mediated lysis (e.g., CD55 and CD59) are GPI-linked, PNH erythrocytes are more sensitive to lysis in vitro by the addition of acidified serum, a source of complement. This somewhat cumbersome assay for PNH (Ham's test) has essentially been replaced in recent years by flow cytometric evaluation of expression of GPI-linked proteins *(31)*. In patients harboring a PNH clone, subsets of erythrocytes, granulocytes, and monocytes demonstrate diminished expression or complete absence of GPI-linked proteins. Such flow cytometric assays are rapid, quantitative, and can be performed even in

patients who have received transfusions, because extremely small populations of GPI-deficient cells are detectable.

REFERENCES

1. Owens, M. A. and Loken M. R. *Flow Cytometry Principles for Clinical Laboratory Practice*. *Quality Assurance for Quantitative Immunophenotyping*. Wiley–Liss, New York, 1995.
2. Ormerod, M. G. (ed.). *Flow Cytometry*. Oxford, New York, 2000.
3. Givan, A. L. *Flow Cytometry First Principles*. Wiley–Liss, New York, 2001.
4. McCoy, J. P., Jr. Basic principles in clinical flow cytometry, in *Flow Cytometry in Clinical Diagnosis* (Keren, D.F., et al. eds.), ASCP Chicago, pp. 31–64, 2001.
5. Leith, C. P. and Willman, C. L. Flow cytometric analysis of hematologic specimens, in *Neoplastic Hematopathology* (Knowles, D. M., ed.), Lippincott Williams & Wilkins, Philadelphia, pp. 255–269, 2001.
6. McCoy, J. P., Jr. Basic principles of flow cytometry. *Hematol. Oncol. Clin. N. Am*. **16:**229–243, 2002.
7. De Rosa, S. C. and Roederer, M. Eleven-color flow cytometry. *Clin. Lab. Med*. **21:**697–712, 2001.
8. Nicholson, J. K. A. and Mandy, F. F. Immunophenotyping in HIV infection, in *Immunophenotyping* (Stewart, C. C. and Nicholson, J. K. A., eds.), Wiley–Liss, New York, pp. 261–289, 2000.
9. Carey, J. L. III, and Keren, D. F. Lymphocyte immunophenotyping and human immunodeficiency virus (HIV) infection, in *Flow Cytometry in Clinical Diagnosis* (Keren, D. F., et al. eds.), ASCP Chicago, pp. 471–505, 2001.
10. Hengel, R. L. and Nicholson, J. K. A. An update on the use of flow cytometry in HIV infection and AIDS. *Clin. Lab. Med*. **21:**841–856, 2001.
11. Schiff, R. and Schiff, S. E. Flow cytometry for primary immunodeficiency diseases, in *Flow Cytometry in Clinical Diagnosis* (Keren, D. F., et al. eds.), ASCP Chicago, pp. 379–470, 2001.
12. Gee, A. P. and Lamb, L. S., Jr. Enumeration of CD34-positive hematopoietic progenitor cells, in *Immunophenotyping* (Stewart, C. C. and Nicholson, J. K. A., eds.), Wiley–Liss, New York, pp. 291–319, 2000.
13. McCoy, J. P., Jr. Enumeration of CD34+ stem cells, in *Flow Cytometry in Clinical Diagnosis* (Keren, D. F., et al. eds.), ASCP Chicago, pp. 581–596, 2001.
14. Jaffe, E. S., et al. (eds.). *Pathology and Genetics of Tumors of Hematopoietic and Lymphoid Tissues*, IARC Lyon, 2001.
15. Stelzer, G. T. and Goodpasture, L. Use of multiparameter flow cytometry and immunophenotyping for the diagnosis and classification of acute myeloid leukemia, in *Immunophenotyping* (Stewart, C. C. and Nicholson, J. K. A., eds.), Wiley–Liss, New York, pp. 215–238, 2000.
16. Leith, C. P. and Carey, J. Flow cytometric analysis of acute leukemia, in *Flow Cytometry in Clinical Diagnosis* (Keren, D. F., et al. eds.), ASCP Chicago, pp. 339–378, 2001.
17. Todd, W. M. Acute myeloid leukemia and related conditions. *Hematol. Oncol. Clin. N. Am*. **16:**301–319, 2002.
18. Borowitz, M. J. and DiGiuseppe, J. A. Acute lymphoblastic leukemia, in *Neoplastic Hematopathology* (Knowles, D. M., ed.), Lippincott Williams & Wilkins, Philadelphia, pp. 1643–1665, 2001.
19. DiGiuseppe, J. A. and Borowitz, M. J. Clinical applications of flow cytometric immunophenotyping in acute lymphoblastic leukemia, in *Immunophenotyping* (Stewart, C. C. and Nicholson, J. K. A., eds.), Wiley–Liss, New York, pp. 161–180, 2000.
20. Riley, R. S., Massey, D., Jackson Cook, C., Idowa, M., and Romagnoli, G. Immunophenotypic analysis of acute lymphocytic leukemia. *Hematol. Oncol. Clin. N. Am*. **16:**245–299, 2002.
21. DiGiuseppe, J. A. and Borowitz, M. J. Clinical utility of flow cytometry in the chronic lymphoid leukemias. *Semin. Oncol.* **25:**6–10, 1998.
22. Carey, J. L. Immunophenotyping of chronic lymphoid leukemias and related non-Hodgkin's leukemias, in *Immunophenotyping* (Stewart, C. C. and Nicholson, J. K. A., eds.), Wiley–Liss, New York, pp. 181–213, 2000.
23. Carey, J. L. Immunophenotyping in diagnosis and prognosis of mature lymphoid leukemias and lymphomas, in *Flow Cytometry in Clinical Diagnosis* (Keren, D. F., et al. eds.), ASCP Chicago, pp. 227–338, 2001.
24. Ben-Ezra, J. B-Cell lymphoproliferative disorders. *Hematol. Oncol. Clin. N. Am*. **16:**321–337, 2002.
25. Witzig, T. E. and Katzmann, J. A. Clinical utility of DNA ploidy and cell proliferation measurements by flow cytometry, in *Flow Cytometry in Clinical Diagnosis* (Keren, D.F., et al. eds.), ASCP Chicago, pp. 663–696, 2001.
26. Crum, C. P. Gestational and placental disorders, in *Pathologic Basis of Disease* (Cotran, R. S., Kumar, V. K. and Collins, T., (eds.), W.B. Saunders, Philadelphia, pp. 1079–1089, 1999.
27. Nguyen, D., Diamond, L. W., and Braylan, R. C. FCM immunophenotyping and DNA analysis: practical aspects that can affect data analysis and interpretation, in *Flow Cytometry in Hematopathology*, Humana Totowa, NJ, pp. 9–36, 2003.
28. Campana, D. and Coustan-Smith, E. Detection of minimal residual disease in acute leuekmia by flow cytometry. *Cytometry* **38:**139–152, 1999.
29. Orfao, A., et al. Residual disease detection of leukemia, in *Immunophenotyping* (Stewart, C.C. and Nicholson, J.K.A., eds.), Wiley–Liss, New York, pp. 239–259, 2000.
30. DiGiuseppe, J. A. and Borowitz, M. J. Leukemias and lymphomas, in *Tumor Markers* (Diamandis, E.P., et al. eds.), AACC Washington, DC, pp. 321–328, 2002.
31. Huard, T., Katzmann, J., and Davis, B. H. Clinically useful nontraditional applications of flow cytometry, in *Flow Cytometry in Clinical Diagnosis* (Keren, D.F., et al. eds.), ASCP Chicago, pp. 543–580, 2001.

14 Medical Cytogenetics

MARTHA B. KEAGLE

1. HISTORY

Medical cytogenetics is the study of human chromosomes under the light microscope and the relationship of chromosomal abnormalities to human diseases, disability, and dysmorphology. As a discipline, human cytogenetics developed slowly at first throughout the 20th centuary. In the early 1900s, the human chromosome number was believed to be 48 for females with an XX sex chromosome complement, and 47 for males with a single X chromosome. In 1928, Painter suggested that both sexes had 48 chromosomes, but that males had two sex chromosomes, an X and a Y *(1)*.

In 1952, Hsu discovered that use of a hypotonic solution swelled cells and separated the chromosomes, making their study much easier *(2)*. This was the first of several technical advances that opened the doors to the clinical evaluation of human chromosomes. Four years later, in 1956, Tijo and Levan, aided by the discovery of hypotonic solution and improved cytogenetic technique, established that the chromosome number in humans was actually 46, not 48 *(3)*.

The first chromosomal abnormality was described by Lejeune in 1959. He discovered that there was an extra copy of chromosome 21 in Down syndrome *(4)*. Later that same year, other researchers reported that Turner syndrome had a 45,X chromosome complement *(5)* and that Klinefelter syndrome had an 47,XXY chromosome complement *(6)*. A woman with 47,XXX was also described *(7)*. The following year, the other common autosomal trisomies (trisomy 13 and 18) were described *(8,9)*. Also in 1960, Nowell and Hungerford described the first consistent chromosomal change seen in cancer. They noticed that in chronic myelogenous leukemia (CML), one chromosome 22 was always unusually small. They dubbed this small chromosome the Philadelphia (Ph) chromosome after the city in which it was discovered. Lejeune described the first deletion syndrome in 1963 and 1964 when he reported that deletion of the short arm of chromosome 5 (cri du chat syndrome) was seen in patients with a catlike cry and phenotypic anomalies *(10,11)*.

Throughout this time, human chromosomes were solid stained using nuclear stains with an affinity for DNA. The classification of individual chromosomes was based on their overall length, the position of the centromere, and the ratio of short to long arms, but the individual chromosomes could not be positively identified and structural rearrangements were difficult, and sometimes impossible, to characterize. In 1970, Caspersson applied to human chromosomes fluorescent techniques that were being used to study plant chromosomes *(12)*. He found that each human chromosome had a unique banding pattern that allowed for the definitive identification of each chromosome pair and made the detection and characterization of structural rearrangements much easier. Over the next several years, a variety of different banding techniques were developed, including ones using Giemsa, introduced by Sumner in 1971 *(13)*, that became fundamental to most cytogenetics laboratories in the United States.

In 1972, Rowley determined that the Ph chromosome seen in CML was actually the result of a reciprocal translocation between a chromosome 22 and a chromosome 9 *(14)*. Prior to banding, the derivative chromosome 9 was not appreciated. In the mid-1970's, Yunis developed techniques of chromosome elongation that allowed chromosomes to be studied earlier in the cell cycle, when they were longer than the typical metaphase preparations that were the mainstay of cytogenetic laboratories *(15)*. Longer prophase and prometaphase chromosomes exhibit more bands than their shorter metaphase counterparts, and this allowed for the detection of smaller structural rearrangements. Pinkel and Gray introduced fluorescence *in situ* hybridization (FISH) in 1986 *(16)*. FISH is a marriage of classical cytogenetics and newer molecular technologies and has extended the range of the former. FISH is dealt with in a chapter of its own elsewhere in this volume.

2. LABORATORY CONSIDERATIONS

Cytogenetics is used to detect gross abnormalities of the human chromosomes that are large enough to be seen with aid of the light microscope. To study human chromosomes, the cells must be dividing, as the individual chromosomes are only visible at this level during mitosis. Routine analyses are performed at metaphase, although techniques to collect cells earlier, in prophase or prometaphase, can be used, as mentioned previously. With the exception of bone marrows, most samples

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

received in the cytogenetics laboratory do not contain actively dividing cells. Therefore, the cells must be grown in tissue culture until sufficient numbers of dividing cells are present.

Cytogenetics laboratories receive four basic types of samples: (1) amniotic fluids, (2) bloods, (3) solid tissues, and (4) bone marrows. Although the individual details differ, the general steps of culture initiation or setup, culture maintenance, and cell harvest are basically the same for all sample types. To obtain dividing cells, the most critical requirement is that there be living cells present in the original sample. It is, therefore, crucial that cytogenetic samples be collected and handled using aseptic technique and that the samples be placed in a sterile growth or transport medium after collection, *not* in formalin or alcohol. Blood and bone marrow samples should be collected in tubes containing sodium heparin to prevent coagulation.

Once received in the laboratory, the cells are placed in sterile tissue culture vessels containing an appropriate complete growth medium. There are many commercial media available that can be used. Some are broad spectrum, suitable for many types of cells; others have been formulated for specific cell types. Microbial inhibitors, including antibiotics and fungicides, can be added to retard microbial growth in the event of contamination.

Blood, bone marrows, and amniotic fluids consist of single cells and can be used as they are received in the laboratory. Lymphocytes from blood and bone marrows are grown in suspension culture in sterile capped centrifuge tubes or T-flasks, whereas fibroblasts from amniotic fluids attach to the inner surface of a T-flask or directly on cover slips in small Petri dishes. The latter is referred to as the *in situ* method of tissue culture. Solid tissues usually must be disaggregated before they are added to the culture medium. This can be done mechanically using sterile scissors or scalpels, or enzymatically. Solid tissues can be grown by the flask or *in situ* method. A few solid tumors grow best in suspension culture.

Whereas most cell types will grow and divide spontaneously in tissue culture if given the proper growth requirements, peripheral lymphocytes must be stimulated to divide. This is done by adding phytohemagglutinin (PHA) to the medium. PHA is a mitogen derived from an extract of kidney beans. There are also B-cell mitogens that could be used when studying hematologic samples from patients suspected of having B-cell disorders. For a more complete discussion of culturing techniques, see refs. *17* and *18*.

After step-up the cultures are placed in an incubator at 37°C and grown until there are sufficient cells in mitosis. The culture time varies depending on sample type. Actively dividing bone marrows cells are usually harvested directly without any time in tissue culture or after 24 h. Routine blood samples are usually allowed to grow for 72 h before harvest, and amniotic fluids can take from several days to a week or longer to produce adequate numbers of dividing cells. Solid tissues often need to culture for a week or longer. When adequate numbers of dividing cells are present, they are harvested.

The mitotic inhibitor, Colcemid, is used to collect dividing cells in metaphase. Colcemid is an analog of colchicine, an extract of the seeds of the autumn crocus. It acts by binding to tubulin and preventing formation of the spindle fibers *(19)*. This prevents separation of the sister chromatids in anaphase, thereby collecting the cells in metaphase, the stage at which they are best observed by light microscopy. The dividing cells are then treated with a hypotonic solution such as potassium chloride. Hypotonic solution swells the cells and aids in the spreading of the chromosomes. Well-spread chromosomes are essential for an accurate cytogenetic evaluation. The cells are then fixed in that swollen state using a modified Carnoy's fixative consisting of three parts methanol to one part glacial acetic acid. This cytogenetic fixative has the added benefit of lysing any red blood cells present in the culture. The harvest procedure is completed by dropping a small amount of the fixed cell suspension onto clean glass microscope slides. There are many variables that influence the spreading and quality of the metaphase preparations. These include ambient temperature, relative humidity and the length of exposure of the cells to hypotonic solution. Good spreading is a function of slide drying, and technologists must be ready with a variety of techniques to compensate for less than optimal chromosome spreading. A longer evaporation time increases spreading, whereas a shorter evaporation time slows spreading. In general, increased temperature and humidity increases chromosome spreading by prolonging the evaporation time, and decreased temperatures and lowered humidity reduces spreading by shortening evaporation time.

The prepared slides are aged and then banded. Aging is essential to achieving good banding. Giemsa-banding (G-banding) is routinely used to study human chromosomes. This technique creates a unique series of light and dark bands along the length of the individual chromosomes (Fig. 1). There are a variety of G-banding techniques (see ref. *18* for sample protocols), but ones involving pretreatment with the enzyme trypsin are the mainstay of most cytogenetics laboratories. Giemsa light bands are rich in the nitrogenous bases CG and are early replicating, whereas the Giemsa dark bands are AT rich and late replicating. A higher proportion of active genes are located in the G-band light regions, making them biologically more significant.

G-banding allows for the positive identification of the individual chromosomes and for the characterization of many structural rearrangements. There are times when G-banding is not sufficient, however. In such cases, there are a wide variety of special stains and FISH probes that can be used to answer specific questions not answered by G-banding. Because a discussion of these is outside the scope of this chapter, the interested reader is referred to refs. *17*, *18*, and *20* for a more complete description of special stains.

3. INDICATIONS FOR CHROMOSOMAL EVALUATION

3.1. AMNIOTIC FLUID The reasons for karyotypic evaluation vary by sample type. Amniotic fluids are the primary sample for prenatal evaluation. Genetic amniocenteses are ideally performed at about 16 wk of gestation and are generally offered to women when their risk of having a chromosomally abnormal fetus is greater than the inherent risk of the amniocentesis procedure itself. Early amniocentesis can be performed at 10–12 wk. The most common reason for performing a genetic amniocentesis is advanced maternal age (AMA), defined as a maternal age of 35 yr or greater at the time of

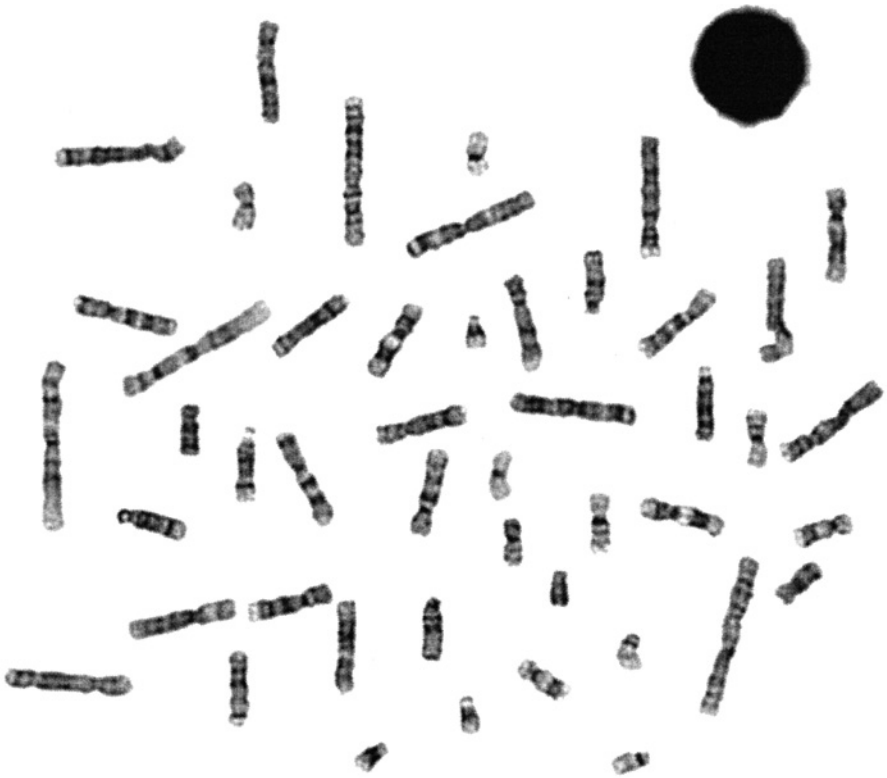


Fig. 1. G-banded metaphase showing the unique banding pattern created along the length of each chromosome.

delivery. Women with AMA are at increased risk of carrying a fetus with a chromosomal abnormality, especially trisomy 21, and are routinely offered an amniocentesis.

A number of maternal serum screening tests such as maternal serum α-fetoprotein, unconjugated estriol, human chorionic gonadotrophin, and dimeric inhibin A have been developed. Women with abnormal values are also at higher risk for certain chromosomal anomalies and, therefore, are candidates for amniocentesis. This is especially important for women under the age of 35 because most cases of Down syndrome occur in this group, but these women are not generally offered amniocentesis on the basis of age alone. Amniocentesis is also indicated when one of the parents is known to be a carrier of a balanced chromosomal rearrangement that could be passed to offspring in an unbalanced form. In addition, this procedure is performed as a follow-up to the finding of fetal anomalies on ultrasound and to clarify possible mosaicism seen in chorionic villus sampling (CVS).

3.2. BLOOD Peripheral blood studies are performed for a large number of reasons. Some of the more common ones are outlined here. Blood studies are often performed on newborns with multiple congenital anomalies. Certain constellations of findings might suggest specific chromosomal syndromes, and ambiguous genitalia might be associated with sex chromosome abnormalities. Chromosomal studies might be indicated for individuals with mental retardation or developmental delay of unknown origin, as they are common findings among individuals with unbalanced chromosomal rearrangements.

Family studies are often indicated once a chromosomal abnormality has been identified in order to rule out carrier status of at risk individuals. The clinical diagnosis of Down syndrome should be confirmed by chromosomal studies because a small portion of cases are the result of familially transmitted translocations and isochromosomes. Therefore, carriers will be at risk for passing on the abnormality.

Primary amenorrhea and premature ovarian failure are indications for chromosomal evaluation. Primary amenorrhea might be caused by Turner syndrome (45,X), a Turner variant or other abnormalities involving the X chromosome. Premature ovarian failure might also be associated with abnormalities involving the X chromosome.

Couples with a history of unexplained multiple pregnancy loss or infertility might harbor a balanced chromosomal aberration that they are passing on to offspring in an unbalanced form and are, therefore, candidates for chromosomal study.

Occasionally, percutaneous umbilical blood sampling (PUBS) is performed for prenatal fetal diagnosis. This could be done as a follow-up to an equivocal amniocentesis result or if fetal anomalies are seen on ultrasound at a later gestational age.

3.3. SOLID TISSUES There are a variety of solid tissues that can be studied cytogenetically. Chorionic villus samples obtained from the chorion frondosum of the developing placenta can be studied for prenatal diagnosis at 8–12 wk gestation, weeks earlier than routine amniocentesis. The earlier results are advantageous for many women in terms of privacy and easier therapeutic termination of pregnancy if one is chosen, but confined placental mosaicism can be problematic in CVS, resulting in equivocal results and the possible need for a follow-up amniocentesis.

Skin biopsies are sometimes performed when a nonhematologic sample is needed or to detect mosaicism that is

Table 1
Common Sex Chromosome Aneuploidies

Turner syndrome	Klinefelter syndrome
Short stature	Tall stature as children
Primary amenorrhea	Tendency toward obesity as adults without testosterone replacement
Failure to develop 2° sex characteristics	Testicular atrophy with azoospermia
Short, webbed neck	Infertility
Low posterior hairline	Gynecomastia
Broad chest with hypoplastic, wide-spaced nipples	Somewhat decreased IQ
Normal verbal IQ, but performance IQs slightly lower than controls	47,XYY
Deficits in visual–spatial organization	Tall Stature
	Large tooth size
47,XXX	No dysmorphology in most; minor skeletal and facial anomalies in some
Normal phenotype	Severe acne
Mental deficiency, psychosis or learning disabilities	Normal fertility
Normal fertility	

tissue-specific, such as seen in Pallister–Killian syndrome. Solid tissue samples can be used to detect chromosomal abnormalities in stillbirths when blood is not available, and products of conception are useful in determining the cause of many spontaneous abortions (SABs). Approximately half of SABs and 5% of stillbirths are chromosomally abnormal *(21)*. Cytogenetic studies can provide valuable information for couples who have experienced pregnancy loss.

Solid tumor tissues can be studied to detect one of the many acquired chromosomal changes associated with malignancy.

3.4. BONE MARROW The vast majority of bone marrow samples are from patients with suspected hematologic disorders for the detection of acquired chromosomal changes. Cytogenetic studies can help establish an initial diagnosis, can be used to detect evidence of relapse after treatment, and can be used to monitor disease progression or recurrence.

4. CONSTITUTIONAL CHROMOSOMAL ABNORMALITIES

Chromosomal abnormalities occur quite commonly. Most are not compatible with life and are aborted spontaneously. Over 50% of early SABs have an abnormal karyotype *(21)*. The attrition rate of fetuses with chromosomal abnormality decreases as pregnancy progresses, and about 5% of stillbirths *(21)* and a smaller percentage of term deliveries are chromosomally abnormal. Chromosomal abnormality is a significant cause of major birth defects. Abnormalities of the sex chromosomes are generally better tolerated than abnormalities of the autosomes and often show a milder phenotype.

Chromosomal abnormalities fall into two broad categories: numerical and structural. Numerical abnormalities include aneuploidy (the presence or absence of entire whole chromosomes) and polyploidy (the presence of entire additional sets of chromosomes).

4.1. ANEUPLOIDY Human chromosomal aneuploidy is seen in about 0.5% of term pregnancies, but it represents about 75% of abnormal karyotypes in SABs *(22)*. In early embryologic development, it has been estimated that 20% of human conceptuses are aneuploid *(23)*. Monosomy is the absence of an entire chromosome, whereas trisomy is the presence of an additional chromosome. Both result from chromosomal nondisjunction during meiosis. Although nondisjunction can occur in either spermatogenesis or oögenesis and in either meiosis I or II, it occurs most commonly in maternal meiosis I. Double trisomies are rare, but they are seen in products of conception.

4.1.1. Monosomy Autosomal monosomies are almost invariably lethal very early in embryonic development and are seldom seen, even in SABs. Rare cases of liveborns with monosomy 21 and 22 have been reported, often in mosaic form *(22)*. Mosaicism is the presence of two or more genetic cell lines. The presence of a normal karyotype in addition to an abnormal one often results in a milder phenotype and, in some cases, the potential for survival. Aneuploid/diploid mosaicism occurs postzygotically as a result of mitotic nondisjunction or from trisomy rescue (loss of one of the extra chromosomes) of a trisomic fetus. Monosomy Y is also lethal; the presence of at least one X chromosome is necessary for survival.

Monosomy X, or Turner syndrome, is the only human monosomy compatible with life. Even so, as many as 99% of cases are aborted spontaneously, and 45,X is the most common karyotype seen in SABs *(21)*. Ironically, the phenotype of Turner syndrome is relatively mild (Table 1) as far a chromosomal abnormalities are concerned. It is postulated that liveborns with a 45,X karyotype could have undetected mosaicism or confined placental mosaicism, which allows their survival.

The 45,X karyotype is seen in somewhat greater than 50% of Turner patients. The remainder have variant karyotypes. Mosaicism involving a 46,XX and/or 47,XXX karyotype or a structurally abnormal X chromosome is common. Structural abnormalities of an X chromosome such as isochromosomes of the long arm, ring chromosomes, and various deletions of the X chromosome are also fairly frequent. Of particular importance are karyotypes involving a Y chromosome, such as 45,X/46,XY mosaics. Females with a Y chromosome or even particular Y sequences are at increased risk of developing gonadoblastomas *(24)*.

Nearly three-quarters of 45,X individuals receive their sole X chromosome from their mother *(25)*. Of interest, those with

a paternally derived X have higher levels of social cognition than those with a maternally derived X. There is apparently a gene for social cognitive functioning that is active on the paternal X, but not on the maternal one *(26)*.

4.1.2. Additional Sex Chromosomes Trisomy of the X chromosome (47,XXX) occurs fairly frequently, being seen in about 1 in 1000 live births. This chromosomal abnormality does not form a well-defined syndrome. Women with this condition are usually phenotypically normal but might be mentally deficient or psychotic or have serious learning deficits. They are fertile and overwhelmingly produce chromosomally normal offspring rather than the one-half XXX females and XXY males that might be expected.

The presence of more than three X chromosomes in females (48,XXXX and 49,XXXXX) is associated mental deficiency and a variety of rather nonspecific physical anomalies. As the number of X chromosomes increases, so does the degree of abnormality. Tetra-X women have menstrual irregularities and reduced fertility, whereas penta-X women appear to be infertile.

Klinefelter syndrome is also present in about 1 in 1000 births. The vast majority of patients have a 47,XXY karyotype, but 46,XX/47,XXY mosaics and Klinefelter variants with a 48,XXYY karyotype also occur. Affected individuals tend to be tall and slim in childhood, but have a tendency toward obesity as adults if not given testosterone replacement therapy. They have testicular atrophy with azoospermia and are infertile. Some develop gynecomastia and have an increased incidence of breast cancer. Intelligence quotients (IQs) are generally somewhat lower than normal, but the range is wide.

As in females, the phenotypic abnormality increases as the number of additional X chromosomes increases in males. Males with 48,XXXY and 49,XXXXY show a variety of physical anomalies and mental deficiency. Infertility is the norm.

A 47,XYY karyotype is seen in about 1 in 1000 male births, but it is usually found incidentally while looking for something else (amniocentesis for advanced maternal age, as part of a family study after detection of a familially transmitted structural abnormality, as part of a cancer study, etc.). This is because most individuals with 47,XYY are not dysmorphic. They tend to have tall stature and large tooth size, and some do show minor anomalies and have a severe type of acne in adolescence. IQs are near normal, although somewhat lower than siblings. Fertility is normal and 47,XYY individuals do not produce an increased incidence of offspring with sex chromosome abnormalities. Several decades ago the question was raised as to whether the presence of an extra Y chromosome predisposed males to violence, criminality, or aggressive behavior. This has not proved to be the case.

4.1.3. Autosomal Aneuploidy Human chromosomal trisomy is very common. Just over one-half of all pregnancy losses with a chromosomal abnormality are trisomic 21. Although trisomy of all chromosomes has been reported in spontaneous abortions, only a few survive to term, and only three of these (trisomies 21, 18, and 13) occur in liveborns with any frequency. Trisomies 8, 9, 14, 15, 21, and 22 have been reported in live births, but usually in mosaic form with a normal cell line. Trisomy 14 has only been reported as a mosaic. Common features of the viable autosomal trisomies are listed in Table 2. Trisomy 16 is the most commonly seen trisomy in spontaneous abortions, but none survive to term (Fig. 2).

Trisomy 21 (Down syndrome) is the most common of all autosomal trisomies to survive to term and is seen in 1 in 700 live births. Even so, 60% of cases of trisomy 21 abort spontaneously. About 94% of cases of trisomy 21 are the sporadic, noninherited type that result from meiotic nondisjunction, whereas 4–5 % result from inherited Robertsonian translocations or isochromosomes of the long arm of chromosome 21, and 1–2 % are mosaic, resulting from mitotic nondisjunction early in embryogenesis. A very small portion of full trisomy 21s are the result of gonadal mosaicism of a parent. This phenomenon accounts for the rare cases of recurrent complete trisomy 21 reported in some families. Mosaicism usually results in a milder phenotype. Life expectancy for individuals with Down syndrome is greater than 45 yr *(27)*.

There are dozens of relatively minor abnormalities seen in trisomy 21 and no patient exhibits all of them. Most of these findings are not particularly abnormal by themselves, but together, they result in a very characteristic appearance. There are a few major abnormalities as well. These include mental retardation, which is fairly universal, heart defects that affect 40–45% of patients, and gastrointestinal abnormalities seen in 10–12% of cases *(27)*.

Trisomy 18 (Edward syndrome) has an incidence of about 1 in 5000–8000 live births. The vast majority of trisomy 18 conceptuses die prenatally, and 90% of liveborns die within the first year of life. A few have survived into their teens. Although most cases are complete trisomy 18 resulting from meiotic nondisjunction, a small percentage are mosaic and typically have a milder phenotype and longer survival. Trisomy 18 is characterized by intrauterine and postnatal growth retardation, profound mental retardation, and a variety of physical anomalies (Table 2).

Trisomy 13 (Pateau syndrome) is seen in 1 in 12,000 live births. Most cases result in SAB and those that continue to term usually have a limited survival. Only 5–10% live to 1-yr and long-term survival is rare. At least 10% of cases result from structural rearrangements such as translocations. Mosaicism does occur, and as with other mosaic conditions, mosaic trisomy 13 can show a broad spectrum of abnormality and longer survival. There appears to be two critical regions on chromosome 13, one proximal and one distal, and both need to be present an extra time for full expression of the syndrome. Profound mental deficiency, midline face defects, cardiac anomalies, and a variety of phenotypic abnormalities (Table 2) are characteristic of trisomy 13.

4.2. POLYPLOIDY Polyploidy can take two forms in humans: triploidy (three sets of chromosomes) and tetraploidy (four sets of chromosomes). Triploidy ($3n = 69$) occurs in about 1% of conceptuses and as many as 20% of spontaneous abortions with a chromosomal abnormality *(27)*. Few triploids survive to term, and those that do usually have a quick demise. Diploid/triploid mosaics have had a longer survival.

Meiotic errors in both oögenesis and spermatogenesis can result in eggs or sperm with an unreduced chromosome number. If normal fertilization occurs, a triploid chromosome complement results. Other possible causes of triploidy include

Table 2
Autosomal Aneuploidies

Trisomy 21 (Down syndrome)
- Mental retardation
- Cardiac anomalies
- GI tract anomalies
- Hypotonia
- Hyperflexible joints
- Tendency to keep mouth open and protrude tongue
- Flat occiput and facial profile
- Upward slant to eyes
- Epicanthal folds
- Brushfield spots
- Small, low-set, abnormally shaped ears
- Short neck with loose folds of skin, especially in infancy
- Fine, soft, sparse hair
- Single palmar crease
- Increased incidence of leukemia
- Decreased incidence of solid tumors, except testicular cancer

Trisomy 18 (Edward syndrome)
- Mental retardation
- Heart defects
- Prominent occiput
- Micrognathia
- Clinched fist and overlapping fingers
- Absent digital crease on fifth finger
- Hypoplastic nails
- Short sternum
- Rocker bottom feet

Trisomy 13 (Pateau syndrome)
- Mental retardation
- Holoprosencephaly
- Cleft lip and palate
- Small eyes or cyclopia
- Coloboma iridium
- Cardiac anomalies
- Polydactyly
- Persistence of fetal hemoglobin
- GU tract abnormalities
- Scalp defects

Trisomy 8 (usually mosaic)
- Mild to severe mental retardation
- Prominent forehead
- Deep-set eyes
- Strabismus
- Thick lips
- Unusual creasing of the palms of hands and soles of feet
- Camptodactyly
- Slender body habitus and narrow pelvis
- Widely spaced nipples

Trisomy 9 (usually mosaic)
- Mental retardation
- Intrauterine and postnatal growth delay
- Microcephaly
- Micrognathia
- Low-set, malformed ears and other facial anomalies
- Joint anomalies

Trisomy 22 (usually mosaic)
- Microcephaly
- Holoprosencephaly
- Hypertelorism
- Hypoplastic, low-set ears
- Limb anomalies
- Genital anomalies

Trisomy 15 (usually mosaic)
- Growth delay
- Developmental delay
- Various craniofacial anomalies
- Limb anomalies
- Clinched fist and overlapping fingers

Trisomy 14 (always mosaic)
- Growth retardation
- Psychomotor retardation
- Central nervous system abnormalities
- Limb abnormalities, including limb asymmetry
- Abnormal skin pigmentation

reunion of a polar body with the egg nucleus or fertilization of a single oöcyte by two sperm. Most triploids have a 69,XXX or 69,XXY karyotype. A 69,XYY karyotype is very uncommon, but it does occur.

The origin of the extra set of chromosomes could be maternal or paternal, and the presentation is very different depending on the origin. If the extra set of chromosomes is paternally derived, the placenta is large and overgrown and the fetus is underdeveloped or undeveloped. If the extra set is maternally derived, the placenta is small and underdeveloped and the fetus is large.

Tetraploidy ($4n = 92$) is much less common and is extremely rare in liveborns, but it is seen in 6–7% of spontaneous abortions with a chromosomal abnormality *(28)*. All reported tetraploids have been 92,XXXX or 92,XXYY. No 92,XXXY or 92,XYYY karyotypes have been described, even in SABs *(28)*.

Most nonmosaic tetraploids probably result from chromosomal duplication without cytoplasmic division in the first postzygotic mitotic division. Other less likely possibilities include the fertilization of a diploid oöcyte by a diploid sperm or fertilization of a normal diploid oöcyte by three sperm's. Diploid/tetraploid mosaicism has been reported with longer survival.

4.3. STRUCTURAL ABNORMALITIES Structural abnormalities occur much less frequently than numerical ones. Whereas numerical abnormalities are always genetically unbalanced, structural rearrangements might be balanced or unbalanced. A balanced rearrangement is one in which all of the genetic material is present; it is just moved around within the karyotype. In unbalanced rearrangements, parts of the genome are represented too few or too many times. Carriers of balanced chromosomal abnormalities are usually phenotypically normal, but they are at risk of producing unbalanced gametes and therefore genetically unbalanced conceptuses, resulting in high rates of spontaneous abortion and/or abnormal offspring.

Translocations (t in cytogenetic nomenclature) are the most commonly occurring structural abnormalities in humans. A

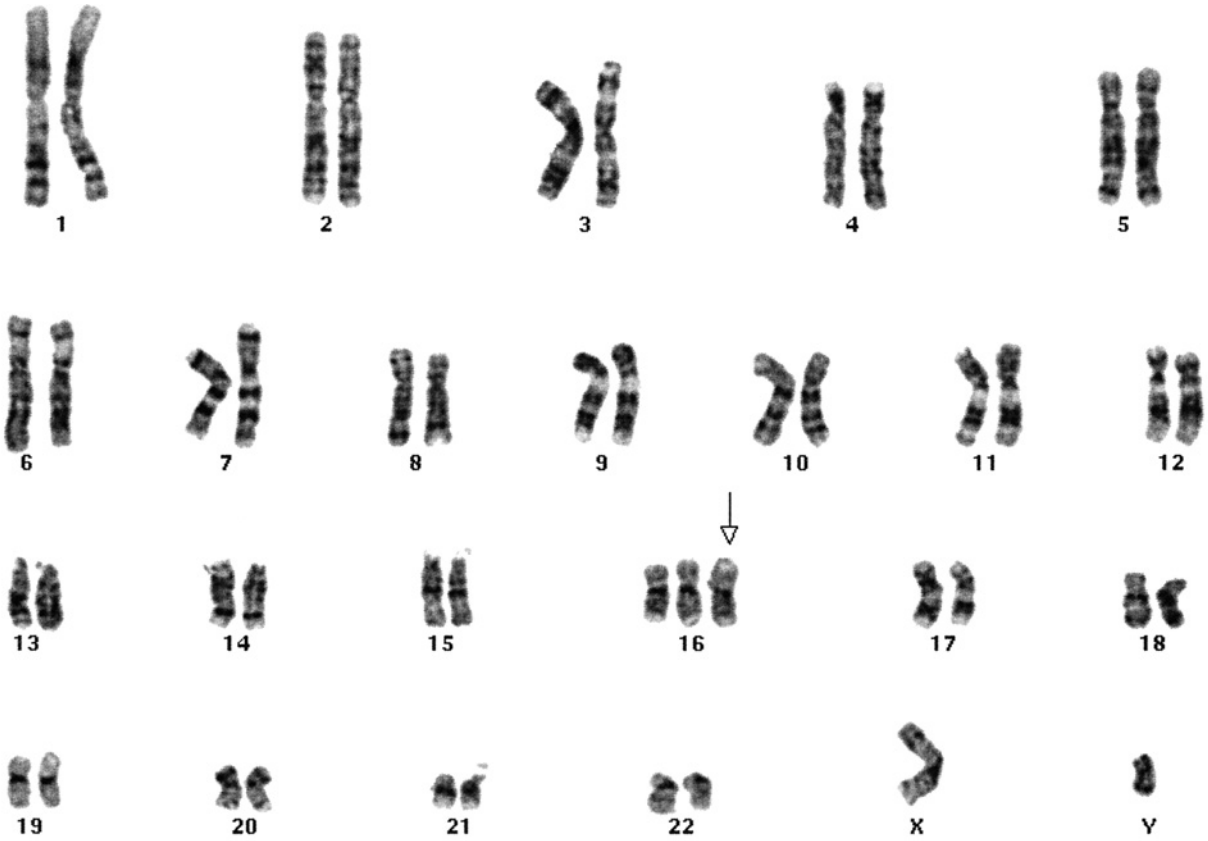


Fig. 2. G-banded male karyotype with trisomy 16. This karyotype is commonly seen in spontaneous abortions.

translocation is an exchange of material between nonhomologous chromosomes. Although balanced translocation carriers are usually phenotypically normal, the translocation chromosomes can malsegregate during meiosis, resulting in genetically unbalanced gametes.

Robertsonian translocations (rob) are a special type of translocation between two acrocentric chromosomes. The result is a single chromosome consisting of the long arms of the two chromosomes involved. This is also referred to as centric fusion. The satellited short-arm material is lost, but this does not result in a genetic imbalance because the short arms of all the acrocentric chromosomes contain redundant copies of the same genetic material. Balanced carriers of Robertsonian translocations have 45 chromosomes instead of the usual 46 but are phenotypically normal. Malsegregation can occur during meiosis, however, and this can result in nullisomic and disomic gametes.

Nonhomologous Robertsonian translocations are the most common recurrent constitutional rearrangements in humans. The most frequently encountered Robertsonian translocations involve chromosomes 13 and 14 and chromosomes 14 and 21.

A deletion (del) is a loss of material from a chromosome. Deletions can be terminal (involving the end of a chromosome arm) or interstitial (involving material within a chromosome arm). Deletions result in the individual being hemizygous for the genes of the deleted region. Not all deletions are visible cytogenetically, even with high- resolution banding techniques. FISH can be helpful in detecting many deletions that are below the level of resolution of light microscopy.

A ring chromosome (r) is a special type of deletion chromosome in which breaks occur in both chromosome arms. The telomeric segments are lost and the broken ends of the piece with the centromere join together to form a donut-shaped chromosome. Ring chromosomes are very unstable and have difficulty accomplishing normal cell division. If crossing-over occurs, the homolog can become tangled or a large dicentric ring could be formed. These might break as cell division progresses. Because of this instability, rings are sometimes lost, giving rise to cells that are monosomic for the chromosome in question.

Duplication chromosomes (dup) contain a chromosomal segment that is represented a second time and that lies adjacent to the original segment. Duplications are referred to as being direct or inverted, depending on the orientation of the duplicated region. If the repeated segment occurs in the same order relative to the centromere as it did in the original chromosome, it is a direct duplication. If the order is reversed relative to the centromere, it is an inverted duplication. Individuals with chromosomal duplications are trisomic for the genes in the duplicated portion.

Inversions (inv) are structural abnormalities involving two breaks in a single chromosome. The region between the breaks rotates 180° and reattaches. If the breaks occur in the same chromosome arm, it is referred to as a paracentric inversion. If the breaks occur in different arms, it is referred to as a pericentric inversion. There are a few inversions that occur frequently enough in the population to be considered polymorphic variants. These are specific inversions of chromosomes 9, Y, and 2.

Table 3
Duplication and Deletion Syndromes

Wolf–Hirschhorn syndrome
- Type of abnormality: deletion
- Critical region: 4p16.3
  - Mental deficiency
  - Ocular hypertelorism
  - Dull expression
  - Strabismus
  - Cleft lip and/or palate
  - Microcephaly and/or cranial asymmetry
  - Turned-down corners of mouth
  - "Greek warrior helmet" face
  - Large, simple, low-set ears with preauricular pits or tags

Cri-du-chat syndrome
- Type of abnormality: deletion
- Critical region: 5p15.2
  - Mental deficiency
  - High-pitched catlike cry in infancy
  - Microcephaly
  - Round, moonlike face
  - Wide-set, downward-slanting eyes
  - Strabismus
  - Epicanthal folds

Williams syndrome
- Type of abnormality: deletion
- Critical region: 7q11.23
  - Mental retardation/developmental delay
  - Broad brow
  - Short, up-turned nose
  - Full cheeks
  - Full lips and wide mouth
  - Long philtrum
  - Stellate pattern to iris
  - Hoarse voice
  - Premature graying and wrinkling
  - Congenital heart defects
  - Hypercalcemia

Langer–Giedion syndrome
- Type of abnormality: deletion
- Critical region: 8q24.11-q24.13
  - Mental retardation
  - Fine, sparse scalp hair
  - Large, bulbous nose
  - Long, simple philtrum
  - Thin upper lip
  - Large, laterally protruding ears
  - Loose, redundant skin in infancy
  - Multiple exostoses
  - Epiphyseal coning
  - Tendency toward bone fractures

Aniridia–Wilms' tumor association (WAGR)
- Type of abnormality: deletion
- Critical region: 11p13
  - Wilms' tumor
  - Aniridia
  - Genitourinary defects
  - Mental retardation

Beckwith–Wiedemann syndrome
- Type of abnormality: Duplication of paternal 11 and other mechanisms
- Critical region: 11p15.5
  - Gigantism
  - Visceromegaly
  - Hemihyperplasia
  - Macroglossia
  - Linear ears creases
  - Hypoglycemia
  - Increased incidence of a number of embryonal tumors (Wilms' tumor, adrenocortical carcinoma, embryonal rhabdomyosarcoma, hepatoblastoma)
  - Pancreatic hyperplasia
  - Abdominal wall defects
  - Fairly normal growth
  - No mental impairment

Pallister–Killian syndrome
- Type of abnormality: mosaic tetrasomy
- Critical region: 12 short arm
  - Mental retardation
  - Sparse hair, eyebrows, and eyelashes
  - Long philtrum
  - Cupid-bow shape to upper lip
  - Protruding lower lip
  - Deafness
  - Minimal speech development
  - Hypotonia
  - Contractures that develop with age

Prader–Willi syndrome
- Type of abnormality: deletion of paternal 15 and other mechanisms
- Critical region: 15p12
  - Mental retardation
  - Early failure to thrive followed by compulsive eating disorder and obesity at about 2 yr
  - Decreased muscle mass
  - Small hands and feet
  - Hypogonadism
  - Almond-shaped eyes
  - Hypopigmentation

Angelman syndrome
- Type of abnormality: deletion of maternal 15 and other mechanisms
- Critical region: 15p12
  - Mental retardation
  - "Marionette like" gait
  - Open mouth and seemingly inappropriate laughter
  - No language development
  - Hypopigmentation

Smith–Magenis syndrome
- Type of abnormality: deletion
- Critical Region: 17p11.2
  - Mental retardation
  - Brachycephaly
  - Low-set, malformed ears
  - Broad nasal bridge
  - Prognathism
  - Cleft lip and/or palate
  - Short fingers
  - Delayed dentition
  - Hoarse voice
  - Sleep disorders
  - Behavior problems

(Continued)

Table 3 *(Continued)*

Hyperactivity
Self-destructive behavior
Attention-seeking behavior

Miller–Dieker syndrome
Type of abnormality: deletion
Critical region: 17p13.3
Lissencephaly
Mental retardation
Microcephaly
Short nose with anteverted nostrils
Vertical ridging of central forehead when crying

Cat–eye syndrome
Type of abnormality: trisomy or tetrasomy of 22q (could be mosaic)
Critical Region: 22q11
Coloboma of the iris
Anal atresia
Down-slanting palpebral fissures
Misshapen ears with preauricular pits or tags
Most have mild mental deficiency

DiGeorge syndrome
Type of abnormality: deletion
Critical region: 22q11.21-11.23
Thymus hypoplasia or aplasia with deficit in cellular immunity
Diminished number of T-cells
Hypoparathyroidism
Cardiac anomalies
Subtle dysmorphic features

Velo-cardio-facial syndrome
Type of abnormality: deletion
Critical region: 22q11.2
Long face
Smooth philtrum
Up-slanting eyes
Cleft lip and/or palate
Long nose with bulbous tip
Small mouth
Slender hands and fingers
Severe hypernasal speech
Learning disorders and psychosis
Congenital heart defects

As with other balanced rearrangements, individuals with inversion chromosomes are usually phenotypically normal, but inverted chromosomes can encounter mechanical problems during meiosis. If the inverted segment is large enough, an inversion loop forms as the chromosomes attempt to pair. If crossing-over occurs within the inversion loop, abnormal gametes with duplicated and deleted chromosomes result.

Insertions (ins) involve the movement of a piece of one chromosome into another or the same chromosome. An insertion requires three breaks: two to create the insertion segment and one in the recipient chromosome. Because of the number of breaks required, insertions are rather rare events.

Isochromosomes (i) are structurally abnormal chromosomes comprised of two identical arms: either two short arms are two long arms. Individuals with isochromosomes will be hemizygous for the genes on the arm missing and trisomic for genes on the arm represented twice.

All of these unbalanced structural abnormalities have their effect by rendering the individual who possesses them genetically imbalanced for specific segments of chromosomes. Some of these are viable and give rise to duplication and deletion syndromes. Although a complete discussion of these is beyond the scope of this chapter, some of the more common duplication and deletion syndromes are described in Table 3. Although most of these syndromes occur sporadically, some are inherited from parents who are carriers of balanced chromosomal abnormalities. Thus, chromosomal study of the biologic parents to rule out carrier status is indicated whenever a structural abnormality is detected in a prenatal sample or in a child.

5. ACQUIRED CHROMOSOMAL ABNORMALITIES

The human genome is inherently unstable with mutations occurring at a low background rate. Most of these changes are lethal and do not survive or they are innocuous and therefore of no harm to the organism, but others are responsible for diseases like cancer. Most cancers result from an accumulation of several mutations. Although some constitutional factors influence the development of cancer, environmental factors such as exposure to carcinogens and ionizing radiation, diet, and certain viruses and bacteria play a major role. Many mutations involved in tumorigenesis occur at the gene level and cannot be detected cytogenetically, but others involve the presence or absence of chromosomes or chromosomal rearrangements that are visible at the level of the light microscope.

Cytogenetically, cancers can be broadly classified as solid tumors and hematologic disorders. Although solid tumors are most important in terms of human morbidity and mortality, the majority of cytogenetic knowledge about cancers deals with the hematologic disorders. Although our knowledge of solid tumors has increased in recent years, it still lags behind that of hematologic malignancies. This is in large part the result of technical difficulties of obtaining a proper sample, difficulties in culturing cells from solid tumors, and because of diagnostic difficulties of interpreting the chromosomal findings, even if decent preparations can be obtained. Because many solid tumors are discovered at a more advanced stage, the cytogenetic picture can be very complex and difficult to decipher.

Flouresence *in situ* hybridization has become a very important tool in cancer genetics in recent years, allowing for rapid and easy detection of specific chromosomal abnormalities known to be common in certain cancers. Despite this, conventional FISH approaches only allow for the detection of the specific abnormalities being probed for, so classical cytogenetics is still important for seeing the larger picture.

Although the cytogenetic changes seen in cancer cells can be very complex and heterogeneous, these acquired changes do not occur randomly throughout the genome. There seems to be a predisposition to breakage at certain sites. Specifically, 83 bands

appear to be of primary importance, and breakage at least 1 of these has been reported to be seen in over 95% of tumor with complex cytogenetic changes *(29).* Cancer genes reside at many of these breakpoints.

There are sometimes particular chromosomal abnormalities that are seen early in tumor development and that might represent an early necessary step in malignant transformation. These are referred to as primary cytogenetic changes and might be seen as the sole abnormality in the tumor cells.

As the tumor develops, additional chromosomal aberrations occur. Many of these are well-documented, nonrandom changes that are never seen as the sole cytogenetic change. They are not necessary for the existence of the tumor, but these secondary changes might give the tumor a growth advantage and represent progression and evolution of the tumor.

Other chromosomal changes appear to be random changes that are unique to a given cell or clone. They might represent unstable mutations that are unable to produce viable cell lines. They are probably not clinically significant and do not appear to play a role in tumor evolution.

Many malignancies show a wide range of chromosomal abnormalities, and many of the most common findings are seen only in a minority of cases. Thus, the abnormalities described must not be considered all inclusive, but they represent some of the more commonly observed changes.

5.1. SOLID TUMORS What follows is just a sampling of the many solid tumors. The reader is referred to other sources for a more complete information on this topic *(30–32).*

5.1.1. Retinoblastoma Retinoblastoma is a malignant neoplasm of the primitive retinal cells of the eye of young children. Retinoblastoma can be either hereditary or sporadic. The *Rb* gene is a tumor suppressor gene located on chromosome 13 at band q14. Loss or mutation of both copies of the gene is required for tumor development. Some individuals have a cytogenetically visible constitutional deletion involving 13q14 that represents the first hit in this two-step process. Other chromosomal changes might also be seen in retinoblastoma. The most commonly occurring cytogenetic change is an isochromosome of the short arm of chromosome 6. This might be seen as a sole change or in conjunction with other cytogenetic changes. Abnormalities of chromosome 1 that result in gains of long-arm material are also seen but have not been reported as the sole cytogenetic change.

5.1.2. Wilms' Tumor Wilms' tumor or nephroblastoma is a genetically complex embryonic kidney tumor affecting children under the age of 7 yr. As with retinoblastoma, there are hereditary and sporadic forms. Loci at 11p13 and 11p15 are important and might be involved in primary cytogenetic changes observed in this tumor. Specifically, deletions of 11p13 and duplications of 11p15 can be seen as primary changes. *WTI*, a tumor suppressor gene, is located at 11p13 and its deletion represents the first hit in tumorigenesis. Several imprinted genes (*IGF2*, *p57*, and *H19)* are located at 11p15 and appear to be involved in development of Wilms' tumor and other cancers. Wilms' tumor might show acquired secondary changes, including whole-arm translocations involving chromosomes 1 and 16 and isochromosomes of the long arm of chromosome 1.

5.1.3. Neuroblastoma Neuroblastoma is one of the small, round, blue cell tumors of childhood. It is a tumor of the neurons of the postganglionic sympathetic nervous system. Deletions involving 1p32-36 are seen as a primary change in as many as 70% of cases. Two possible tumor suppressor genes, *NB-R1* and *NB-R2*, are located in this region and appear to play a role in tumorigenesis. Tumors with deletions of *NB-R1* are usually triploid, whereas those with deletions of *NB-R2* are diploid or tetraploid. There is a characteristic translocation, t(1;17)(p36;q12), seen as a secondary change in neuroblastoma, and gene amplification is seen in about half of the cases. Amplification of the *N-myc* oncogene located at 2p24.1 results in double minutes (dmin) or, less frequently, homogeneously staining regions (hsr).

Double minutes are small, double structures in the cell nucleus that represent amplified genes that have been moved from their chromosome of origin into the nuclear matrix. Double minutes can vary in size and number from cell to cell. The variability in number results from their unequal distribution following cell division. This occurs because double minutes do not contain centromeres or kinetochores and have no spindle attachment.

Homogeneously staining regions are relatively uniformly staining segments that represent incorporation of amplified genes on recognizable chromosomes. In most cancers, double minutes and homogeneously staining regions are associated with advanced disease and poor prognosis. In neuroblastoma, however, although not a primary change, gene amplification is not a late change and is usually present at the time of diagnosis.

5.1.4. Bladder Cancer Eighty-five to 95% of bladder cancers in the United States are transitional cell carcinomas arising from the transitional epithelium lining the bladder. Common primary changes in transitional cell carcinoma include trisomy 7, deletions involving 21q22, and deletions involving 10q22-24. The latter is the most commonly occurring structural change seen in transitional cell carcinoma.

Secondary cytogenetic changes include isochromosomes of the short arms of chromosomes 5 and 11 and the long arm of chromosome 11, deletions and translocations of 11p11–q11, deletions and translocations involving 1q21, and monosomy 9. Monosomy 9 is the most common cytogenetic change seen in transitional cell carcinoma. Deletions and rearrangements involving 13p14 are also seen and loss or inactivation of the *Rb* gene might be involved.

5.1.5. Small Cell Carcinoma of the Lung Of the four major types of lung cancer (adenocarcinoma, squamous cell carcinoma, small cell carcinoma and large-cell undifferentiated carcinoma), small cell lung cancer (SCLC) was the first to be characterized cytogenetically. SCLC is one of the most highly metastatic cancers in man.

A specific deletion of chromosome 3p14-p23 is seen in almost all cases of SCLC and is very specific to it. Trisomy 7 is also seen as a primary cytogenetic change. Trisomy 7 is not tumor-specific and is seen in many different tumors. It might not even be a malignancy-associated change, as it has been seen in benign cells around the lung cancer. Nevertheless, an extra chromosome 7 is often seen early in small cell cancer.

A variety of secondary changes might also be seen in SCLC. Most result in loss of heterozygosity of particular gene loci where known or suspected tumor suppressor genes reside. These include deletions of the long arms of chromosomes 5 (*APC* is located at 15q21-22) and 13 (*Rb* is located at 13q14), deletions of the short arms of chromosomes 9 and 17 (*p53* is located at 17p13.1), and various deletions of chromosome 6. Gene amplification is also seen as a secondary change. These often involve the myc family genes and can take the form of double minutes or homogeneously staining regions. They are a late manifestation and are associated with poor prognosis.

5.1.6. Breast Cancer Breast cancer is very common in the United States and is the second leading cause of cancer deaths in US women. Primary cytogenetic changes include isochromosomes of the long arm of chromosome 1 and an unbalanced whole-arm translocation between chromosome 1 and 16. This results in the cells having two normal chromosomes 1, one normal chromosome 16, and a fourth chromosome consisting of the long arm of chromosome 1 and the short arm of chromosome 16. Monosomy 17 is another primary change seen in breast cancer. There are several cancer genes on chromosome 17 and loss of heterozygosity of chromosome 17 seems to play a role in some breast cancers.

There a large number of secondary changes that can be seen in breast cancer, including deletions of the short arm of chromosome 3, deletions of 6q21-22, trisomies of chromosomes 7, 18, and 20, and various rearrangements of chromosomes 1, 3, 7, and 11. Gene amplification is fairly common in breast cancer. It presents most frequently as double minutes, but homogeneously staining regions also occur. Gene amplification in breast cancer is more common in metastatic disease. Amplification of *ERBB2,* c-*myc*, and amplicons mapped to 11q13 and 20q13 have all been reported. A homogeneously staining region on the short arm of chromosome 8 is another recurrent example of gene amplification in breast cancer.

5.1.7. Colorectal Cancer Colorectal cancer is one of the leading causes of human morbidity and mortality. It is the fourth most frequently occurring cancer in the United States and the second leading cause of cancer deaths, surpassed only by lung cancer. The cytogenetics of colorectal cancers falls into two distinct categories: those characterized by polyploidy (specifically near-triploidy and near-tetraploidy) and those characterized by near-diploidy or pseudodiploidy (a chromosome count of 46 but with missing and extra chromosomes).

Those tumors that are near-triploid or near-tetraploid have complex karyotypes involving both numerical and structural abnormalities. These tumors tend to be poorly differentiated and are associated with short survival. By contrast, those that are near-diploid or pseudodiploid have simpler karyotypes involving primarily numerical abnormalities and unrelated clones. These tumors tend to be moderately to well differentiated and patient survival is longer.

Gains of chromosome 7 and 20 seem to be primary changes in colorectal cancer. An additional chromosome 7 has also been reported in benign adenomas of the colon, so although an early chromosomal change, it might not be a malignancy associated change *per se.*

A large number of recurrent secondary changes have been seen in colorectal cancer. These include isochromosomes of the long arms of chromosomes 1, 7, 8, 13, and 17. Isochromosome 17q results in loss of the *p53* gene. Deletions of 1q13, the long arm of 12, and 17p11 are also reported. The latter again results in the loss of *p53*. Trisomies of chromosomes 8 and 12 also occur. Trisomy 12 results in their being an extra copy of the *K-ras* oncogene. Monosomy 18 results in loss of heterozygosity of *DCC* and other suppressor genes on 18q. Double minutes are seen in many colorectal cancers and are generally seen late in disease progression and carry a poor prognosis.

5.2. HEMATOLOGIC MALIGNANCIES Hematologic malignancies could be of myeloid or lymphoid origin. Those of myeloid origin include the myeloproliferative disorders (MPD's), the myelodysplastic syndromes (MDS's), and the acute myelogenous leukemias (AML's). Those of lymphoid origin include acute lymphocytic leukemia (ALL), the chronic lymphoproliferative disorders, and the lymphomas.

5.2.1. Myeloproliferative Disorders These are disorders that arise at the level of the pluripotent stem cell. Thus, all myeloid lineages are involved, although one predominates. There is overproduction of the three major cell lines in the bone marrow, giving rise to leukocytosis, thrombocytosis, and erythrocytosis. The MPDs include chronic myelogenous leukemia (CML), essential thrombocytopenia (ET), polycythemia vera (PV) and myelofibrosis with myeloid metaplasia (MF/MM).

Chronic myelogenous leukemia (CML) represents 15–25% of all leukemias. It can occur in all age groups, but is most common in the fifth decade of life. CML is characterized by involvement of all three cell lineages, but overproduction of granulocytes is pronounced. The disease is characterized by three phases: (1) chronic, (2) accelerated, and (3) acute.

The chronic phase is relatively benign and lasts 3–5 yr. There is unregulated proliferation of the white cells, resulting in elevated white blood cell counts, but patients are often deceptively healthy in appearance. Fatigue, weight loss, and splenomegaly are sometimes seen. In the chronic phase, blood smears reveal a mixture of cell types. In the accelerated phase, the cells begin to lose their ability to differentiate and the peripheral white cell counts become even higher. In most cases, the disease enters an acute phase. This transformation can be gradual or rapid as the patient enters what is referred to as blast crisis. Increased numbers of immature cells, blasts, are seen in the bone marrow and usually the peripheral circulation. CML most often transforms to acute myelogenous leukemia (AML), subtypes M1, M2, and M4.

The primary cytogenetic change in CML is a balanced translocation between chromosomes 9 and 22 (t(9;22)(q34.1;q11.2)), the so-called Philadelphia translocation. This translocation brings together the *c-ABL* gene on chromosome 9 and BCR on chromosome 22 creating a hybrid gene the BCR-ABL fusion gene that is responsible for the disease. Most cases of CML will show the standard t(9;22) cytogenetically, but 10–15% will show a variant Philadelphia translocation involving additional chromosomes. Some cases show no cytogenetically visible translocation at all, although the gene rearrangement has taken place at the molecular level. For this reason, molecular techniques are considered the gold standard for diagnosing the

Philadelphia rearrangement. Cytogenetics still has a place in the management of patients with CML, however, as it will pick up secondary changes indicative of disease progression.

The Philadelphia rearrangement occurs early in tumor development and is the sole cytogenetic change early in the disease course. As the disease progresses, especially during the accelerated phase and blast crisis, additional chromosomal abnormalities appear and give rise to increasingly malignant subclones. The most frequently occurring secondary changes are the presence of a second Philadelphia chromosome, an extra chromosome 19, an extra chromosome 8, and an isochromosome of the long arm of chromosome 17. These additional changes can occur singly or in various combinations.

The other myeloproliferative disorders occur in older adults, generally after the age of 40. Essential thrombocytopenia is characterized by a marked increase of platelets in the peripheral circulation and megakaryocytes in the bone marrow. There is an overproduction of erythrocytes in polycythemia vera, and fibrosis of the bone marrow and extramedullary hematopoiesis characterize myelofibrosis with myeloid metaplasia. Fibrosis in the latter can make it difficult to obtain an adequate bone marrow sample.

Unlike CML, there are no specific chromosomal changes associated with the other myeloproliferative disorders. There are a variety of nonspecific changes that are seen with some frequency, however. These include deletions of the long arms of chromosome 13 and 20, monosomy 7, and trisomy 8 and 9.

5.2.2. Myelodysplastic Syndromes Refractory anemia (RA), RA with ringed sideroblasts (RARS), RA with excess blasts (RAEB), RAEB in transformation (REAB-T), and chronic myelomonocytic leukemia (CMML) comprise the myelodysplastic syndromes. All three major bone marrow cell lines are involved in the myelodysplastic sundromes (MDSs) because the malignant event occurs in the pluripotent stem cell or the myeloid progenitor cell. The MDSs exhibit hypercellularity of the bone marrow but lowered cell counts in the peripheral circulation. This occurs because there is an overproduction of cellular elements in the bone marrow, but their maturation is abnormal or dysplastic, and ineffective. Therefore patients, exhibit anemia, leucopenia, and/or thrombocytopenia. CMML leukemia is the exception to this. Patients with CMML have increased white blood cells and monocytes in both their bone marrow and peripheral circulation. The MDSs can evolve into acute leukemia.

Acquired cytogenetic abnormalities are seen in about 70% of cases of MDS. There are no chromosomal abnormalities specific to any of the MDSs, but cytogenetics can be useful in establishing the presence of a malignancy and differentiating it from benign conditions that also present with anemia, leukopenia, and/or thrombocytopenia.

Deletions of the long arm of chromosome 5 are seen frequently in the MDSs. They could be seen as a sole change and generally carry a poor prognosis. Some of the other cytogenetic abnormalities seen in MDS include deletions of the long arm of chromosome 7 and monosomy of chromosome 5 and 7. Many other changes have also been reported.

5.2.3. Acute Myelogenous Leukemia Acute myelogenous leukemia can occur in any age group but is most common under one year and over 30 yrs of age. There is a marked increase after age 55–60. AML is characterized by a malignant accumulation of immature blast cells that replaces the normal cells of the bone marrow. Replacement of the normal cellular elements results in anemia, which causes fatigue and malaise, infections and fever, and bruising and hemorrhage. Patients with AML are usually sick when they present with the disease. The increased numbers of abnormal cells seen in CML are not the result of a proliferation but rather to a block in maturation of the cells and reduced rates of cell death.

The French–American–British (FAB) classification is one of many classification systems for AML. It categorizes AML into eight subtypes, M0–M7, based on cellular morphology and the differentiation pathway involved. Although the cytogenetic findings in some of the subgroups is nonspecific, it is very specific in others, and some findings have important prognostic value.

M0: Acute myelogenous leukemia without maturation. This is a malignancy of primitive blasts. There is no maturation of the cells and their origin as myeloblasts cannot be determined without use of electron microscopy or immunophenotyping. M0 represents less than 5% of AML cases. A variety of numerical and structural abnormalities can be seen in M0, but none is specific to it. A translocation, t(1;11)(p32;q23), has been reported in M0 and M5, and an inversion of the long arm of chromosome 3, inv(3) (q21q26), has been seen in M0 as well as most other AML subtypes.

M1: Acute myelogenous leukemia with minimal maturation. Blasts predominate in M1, but Auer rods might be seen. Auer rods are a fusion of cytoplasmic granules and are pathognomonic for cells of myeloid origin. This subtype accounts for about 20% of cases of AML. The cytogenetic findings of M1 are nonspecific and include trisomy 8, monosomy 7 and the inv(3)(q21q26) mentioned earlier. A translocation, t(3;3)(q21;q26), has been reported in M1 and several other AML subtypes.

M2: Acute myelogenous leukemia with maturation. This is the most common subtype of AML accounting for 30% of cases. Although immature cells prevail, Auer rods might be present and there is more maturation than in M1. A characteristic translocation, t(8;21)(q22;q22), is seen in 40% of cases of M2, and in 15% as the sole cytogenetic change. The translocation brings together *ETO* on chromosome 8 and *AML1* on chromosome 21. Other cytogenetic changes include trisomy 8, trisomy 4, monosomy 7, and loss of the Y chromosome. The latter is seen as a secondary change, especially in men with the 8;21 translocation.

M3: Acute promyelocytic leukemia (APL). M3 is characterized by the presence of numerous primary granules and Auer rods. There is a hypergranular form in which the granules are very pronounced and a microgranular form in which the granules are less prominent. There is also a M3v variant in which the granules are not visible under light microscopy. Ten percent of AMLs are the M3 subtype. This is a slow-growing tumor. Direct cultures will often not reveal chromosomal abnormalities, so 24- and 48-h cultures are advised if APL is suspected.

Acute promyelocytic leukemia (APL) is characterized by a very specific translocation, t(15;17)(q22;q21.1), that is seen in over 90% of cases, but in no other cancer. The *RARα* gene that codes for retinoic acid is located at 17q21.1 and this translocation brings it together with the *PML* gene located at 15q22, creating the *PML/RARα* fusion gene.

Chemotherapy alone cannot be used to treat APL because it lyses the promyelocytes. This releases the cell contents, which disrupts coagulation and can result in fatal diffuse bleeding. To overcome this problem, all-*trans*-retinoic acid (ATRA) is administered prior to the chemotherapy. ATRA is a vitamin A analog that matures the promyelocytes to polymorphonuclear leukocytes, thus avoiding the potentially fatal sequela.

M4: Acute myelomonocytic leukemia (AMML). This leukemia is a mixture of monocytic and granulocytic cells. There is a variant, M4 with eosinophilia (M4eo), in which eosinophils predominate. M4 represents about 25% of AMLs. An inversion, inv(16)(p13q22), is seen in 30% of M4eo and occasionally with other subtypes of AML. The inversion brings together *MYH11* at 16p13 and *CBFβ* at 16q22, creating the *CBFβ/MYH11* fusion gene that has transforming ability. This is a clinically significant finding, as patients with the inverted 16 have a good prognosis, especially when treated with cytosine–arabinoside. Other cytogenetic changes include a deletion of 16q22, also seen in M4eo, trisomy 8, and monosomy 7. An additional chromosome 22 is sometimes seen in conjunction with the inverted 16.

M5: Acute monoblastic leukemia (AMoL). About 10% of AMLs are M5. M5a is characterized greater by than 80% immature monoblasts, whereas M5b has greater than 20% more mature monoblasts. A variety of structural abnormalities involving the 11q23 breakpoint have been described in M5, including t(9;11)(p22;q23), t(11;17)(q23;q21), t(11;17)(q23;q23), t(11;19)(q23;p13.3), and t(11;19)(q23;p13.1). The *MLL* gene is located at 11q23 and appears to be involved in the creation of fusion genes in the various rearrangements. The t(9;11) is subtle, involving an exchange of very similar segments and could therefore be under recognized. The 9;11 translocation occurs most frequently in the M5a subtype. Trisomy 8 is also seen in M5.

M6: Acute erythroleukemia (AEL). Fewer than 5% of AMLs are the M6 subtype. The neoplasm consists of myeloblasts and erythroblasts. The inversion, inv(3)(q21q26), has been reported in M6 and most other AML subtypes. A translocation, t(3;5)(q25;q34), and the nonspecific finding of monosomy 7 and trisomy 8 has also been reported in M6.

M7: Acute megakaryocytic leukemia. M7 accounts for less than 5% of all AMLs. It is characterized by megakaryocytes and myelofibrosis. As with other conditions involving myelofibrosis, the fibrosis can make it difficult to obtain a good cytogenetic sample. A number of structural rearrangements have been reported in M7, but most are nonspecific and are seen in a number of other AML subtypes as well. The exception is a translocation, t(1;22)(p13;q13). This rearrangement is seen especially in infants and children and is the most common abnormality seen in young children with M7.

5.2.4. Acute Lymphocytic Leukemia Acute lymphocytic leukemia (ALL) is a malignant proliferation of immature lymphoid in the bone marrow and usually the peripheral circulation. Peripheral blood studies reveal anemia, thrombocytopenia, and usually increased numbers of lymphoblasts. Most cases are of B-cell origin, so use of a B-cell mitogen is recommended by some.

Acute lymphocytic leukemia can occur at any age, but is seen most often in children age 1–10 yr, with most cases occurring between 3 and 5 yr. Symptoms of ALL include fatigue, bruising, hemorrhage, bone pain (especially in younger patients), and neurologic symptoms if there has been involvement of the central nervous system.

Chromosomes from ALL often exhibit poor quality and poor staining, making them diagnostically challenging. Nevertheless, about two-thirds of cases of ALL demonstrate recognizable cytogenetic changes, and the cytogenetic findings often have prognostic significance.

Hyperdiploidy is seen in 25–30% of childhood ALL, but only 10% of adult ALL. Massive hyperdiploidy (a chromosome number of greater than 50), and especially the occurrence of 54–57 chromosomes without structural anomalies, carries a good prognosis and has the best prognosis of all cytogenetic abnormalities in ALL. Hyperdiploidy of 47–50 chromosomes and pseudodiploidy have an intermediate prognosis.

Near-haploidy (a chromosome number of less than 30 and usually 23–28) is associated with a generally poor outcome. It is important not to overlook apparently broken metaphase spreads when performing the cytogenetic evaluation, as they could represent near-haploid cells.

A variety of structural abnormalities could be seen in ALL. Almost all carry a poor prognosis. A translocation, t(9;22)(q34.1;q11), looks cytogenetically like the one seen in CML, but the breakpoint on the chromosome 22 is slightly different at the molecular level. This translocation is associated with an extremely poor prognosis and is seen more frequently as patient age increases. A translocation, t(4;11)(q21;q23), is seen in infants with ALL and also carries a poor prognosis. When seen in older patients, it is usually associated with prior exposure to genotoxic agents. Rearrangements involving 11q23 are common in treatment-related leukemias. In this translocation, *MLL*, located at 11q23, and *AF4*, located at 4q21, are fused, producing a new transcription factor.

There is one rearrangement seen in pediatric ALL that carries a good prognosis. This is the translocation t(12;21)(p13;q22). It creates a fusion between *AML1* at 21q22 and *TEL* at 12p13. This translocation cannot be visualized cytogenetically, but it can be identified with FISH probes that detect the fusion of these two genes.

Translocations involving a 14q11 breakpoint are seen with some frequency in T-cell ALL. Some examples are t(10;14)(q24;q11), t(11;14)(p13-15;q11), and t(8;14)(q24;q11). The T-cell receptor-α (TCRA) locus resides at 14q11 and it appears that juxtaposition of it with a variety of oncogenes activates those oncogenes. In general, T-cell ALL patients have a poorer prognosis than patients with B-cell ALL.

5.2.5. Chronic Lymphoproliferative Disorders Of the many chronic lymphoproliferative disorders, only two, chronic

lymphocytic leukemia (CLL) and multiple myeloma (MM), will be addressed here.

Chronic lymphocytic leukemia represents about 30% of all leukemias in the United States and Europe and is the most commonly occurring leukemia in these populations. CLL is a disease of middle-aged and elderly adults. It is rare before the fifth decade. Many patients are asymptomatic when they are diagnosed with the disease, but others present with lymphadenopathy or splenomegaly. The peripheral blood shows increased numbers of white blood cells with mature lymphocytes predominating. For this reason, peripheral blood samples are often adequate for detecting chromosomal abnormalities associated with CLL. CLL has a low mitotic rate, so culture intervals of 5–7 d are generally recommended rather than the standard direct or 24-hr cultures that give best result for most leukemias.

Over 95% of CLLs are of B-cell origin, so use of a B-cell mitogen is generally recommended. Although T-cell CLL is relatively uncommon, patients with the chromosome instability syndrome ataxia telangiectasia have an increased incidence of it. B-Cell CLL has a good prognosis, whereas T-cell CLL carries a poor prognosis. CLL can transform, becoming an acute leukemia, most often ALL.

Trisomy 12 is the most commonly encountered cytogenetic change in B-cell CLL, being seen in about one-third of cases. The translocation t(11;14)(q13;q32) is the most common translocation seen in B-cell CLL, and other rearrangements involving chromosome 14 also occur. Deletions and translocations involving the long arm of chromosome 13 are seen fairly frequently.

The inversion inv(14)(q11q32) is commonly seen in T-cell CLL as are other rearrangements involving q11 and q32 breakpoints on chromosome 14. As mentioned previously, the TCRA locus is located at 14q11.

Multiple myeloma is a B-cell tumor in which there is a malignant proliferation of plasma cells that impairs bone marrow function. The abnormal cells produce monoclonal immunoglobulin with excess light chains. MM is also characterized by lytic bone lesions and impaired renal function resulting from excretion of the surplus light chains (Bence–Jones proteins). Patients with MM might bone pain and fractures, anemia and fatigue, and excessive urination and thirst. X-rays might demonstrate the lytic bone lesions.

Plasma cells divide slowly. As a result, the majority of cells in cytogenetic preparations might represent normal bone marrow components. Obviously, these will not show chromosomal abnormalities that might be present in the abnormal plasma cells. Analysis of large numbers of cells is often necessary to improve the likelihood that the appropriate cells are being detected. FISH can be helpful by allowing large numbers of cells to be screened.

Thirty to fifty percent of cases show cytogenetic changes, and of these, one-quarter involve abnormalities of chromosome 14. A third of these are a translocation, t(11;14)(q13;q32). The *IgH* gene is located at 14q32.22 and *BCL1* is located at 11q13. The juxtaposition of the two genes activates *BCL1*. Twenty percent of cases with a chromosomal abnormality show rearrangements of chromosome 11. Half of these are the t(11;14) just mentioned. This translocation is generally seen with monosomy 13, deletions of the long arm of chromosome 13, or abnormalities of chromosome 1. Abnormalities of chromosome 1 are seem in 40% of cases with a cytogenetic change. Deletions of the long arm of chromosome 6 are seen in 15% of cases and hyperdiploidy is seen in two-thirds of chromosomally abnormal cases.

5.2.6. Lymphomas Lymphomas can be broadly classifies as Hodgkin lymphoma and non-Hodgkin lymphoma (NHL). Hodgkin lymphoma is the most common lymphoma. It occurs in young people and is seen more frequently in males than females. Patients often present with lymph node enlargement, but might also experience fever or weight loss.

A large binucleated cell, the Reed–Sternberg cell, characterizes Hodgkin lymphoma cytologically. This is thought to be the malignant cells of Hodgkin lymphoma, but only comprises about 5% of the tumor. The majority of the tumor, 95%, consists of benign histiocytes, lymphocytes, eosinophils, and plasma cells. Because of the large numbers of benign cells present in this tumor, a standard cytogenetic study could reveal few if any malignant cells. For this reason, a large number of cells must be evaluated to provide meaningful information.

Many structural abnormalities have been seen in Hodgkin lymphoma, with rearrangements involving chromosome 1 being seen in about a third of cases. Other anomalies include additional material on the long arm of chromosome 14, deletions of the long arm of chromosome 6, isochromosomes of the short arm of chromosome 6, and abnormalities involving the long arms of chromosomes 3 and 7 and the short arms of chromosomes 12 and 13. Trisomies of chromosomes 1, 3, 7, 8, and 21 are the most commonly occurring numerical abnormalities.

Non Hodgkin lymphoma is a complex group of neoplasms that over the years has been classified in numerous ways. Cytogenetic abnormalities are seen in 90% of cases, far greater than in any leukemia. The cytogenetics is often very complex, and this, combined with the confusing classification of NHL, has made comprehension of the karyotypic findings difficult. One NHL, Burkitt's lymphoma, does show highly specific and consist cytogenetic changes, however.

Burkitt lymphoma is a an immunoglobulin (Ig) producing, B-cell tumor. The Epstein–Barr virus might play a role in its development. Burkitt's lymphoma occurs endemically in Africa and nonendemically in other parts of the world.

Eighty percent of cases of Burkitt lymphoma show a specific translocation, t(8;14)(q24;q32). The remaining 20% show one of two variant translocations. Two thirds of these show t(8;22)(q24;q11) and the remaining one-third show t(2;8)(p11;q24). What these translocations have in common is an 8q24 breakpoint. The c-*myc* oncogene is located at this breakpoint and the genes for the Ig heavy chain, kappa light chain, and lambda light chain reside at 14q32, 2q24, and 22q11, respectively. c-*myc* translation appears to be activated by the juxtaposition of the oncogene next to the Ig heavy, kappa light chain, or lambda light chain genes.

A number of other nonrandom structural and numerical abnormalities have been observed in Burkitt lymphoma. Secondary changes include abnormalities of the short arm of chromosome 1, seen in 30% of cases, and abnormalities involving a 13q34 breakpoint, seen in 15% of cases.

5.2.7. Treatment-Related Hematologic Disorders Over the years, there has been prolonged survival and even cure of cancer patients because of better methods of treatment, especially involving radiation and chemotherapy. Sometimes, these treatments induce new malignancies, often treatment-related myelodysplastic disorders and AMLs (t-MDS and t-AML).

Hypodiploidy (fewer than 46 chromosomes) is seen in over half of such cases. Monosomy 7 is seen in half of cases and monosomy 5 is present in 20–25% of cases. Deletions of the long arm of chromosomes 7 and 5 are also fairly common. Patients who have received topoisomerase II inhibitors as part of their treatment frequently show balanced rearrangements involving 11q23 and 21q22 (*MLL* and *AML1* are located at these breakpoints, respectively), the inverted 16 characteristic of AML M4eo and the translocation t(15;17), typical of AML M3.

ACKNOWLEDGMENT

My thanks to Judy Brown for providing the G-banded metaphase and karyotype.

REFERENCES

1. Painter, T. S. Studies in mammalian spermatogenesis. II The spermatogenesis in man. *J. Exp. Zool.* **37:**291–336, 1923.
2. Hsu, T. S. Mammalian chromosomes in vitro. I The karyotype in man. *J. Hered.* **43:**167–172, 1952
3. Tijo, H. J. and Levan, A. The chromosome numbers of man. *Hereditas* **42:**1–6, 1956.
4. Lejeune, J., Gautier, M., and Turpin, R. Étude des chromosomes somatiques de neuf enfants mongoliens. *Contemp. Rend. Acad. Sci.* **248:**1721–1722, 1959
5. Ford, C. E., Miller, O. J., Polani P. E., Almeida, J. C. de., and Briggs, J. H. A sex-chromosome anomaly in a case of gonadal dysgenesis (Turner's syndrome). *Lancet* **I:**711–713.
6. Jacobs, P. A. and Strong, J. A. A case of human intersexuality having a possible XXY sex-determining mechanism. *Nature* **183:**302–303, 1959.
7. Jacobs, P. A., Baikie. A. G., MacGregor, T. N., and Harnden, D. G. Evidence for the existence of the human "superfemale." *Lancet* **II:**423–425, 1959.
8. Patau, K., Smith, D. W., Therman. E., and Inhorn, S. L. Multiple congenital anomaly caused by an extra chromosome. *Lancet* **I:**790–793, 1960.
9. Edwards, J. H., Harnden, D. G., Cameron, A. H., Cross, V. M., and Wolff, O. H. A new trisomic syndrome. *Lancet* **I:**711–713, 1960.
10. Lejeune, J., Lafourcade, J., Berger, R., et al. Trois cas de délétion partielle du bras court d'un chromosome 5. *C.R. Acad. Sci. (Paris)* **257:**3098–3102, 1963.
11. Lejeune, J., Lafourcade, J., de Grouchy, J., et al. Délétion partielle du bras court un chromosome 5. Individualisation d'un nouvel état morbide. *Sem. Hôp. Paris* **18:**1069–1079, 1964.
12. Nowell, P. C. and Hungerford, D. A. A minute chromosome in human chronic granulocytic leukemia. *Science* **132:**1497, 1960.
13. Sumner, A. T., Evans, H. J., and Buckland, R. A., New technique for distinguishing between human chromosomes. *Nature New Biol.* **232:**31–32,1971.
14. Rowley, J. D. A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **243:**290–293,1973.
15. Yunis, J. High resolution of human chromosomes. *Science* **191:**1268–1270, 1976.
16. Pinkel, D., Gray, J., Trask, B., van den Engh, G., Fusco, J., and van Dekken, H. Cytogenetic analysis by in situ hybridization with fluorescently labeled nucleic acid probes. *Cold Spring Harbor Symp. Quant. Biol.* **51:**151–157, 1986.
17. Keagle, M. B. and Gersen, S. L. Basic laboratory Procedures, in *The Principles of Clinical Cytogenetics,* Gersen, S.L. and Keagle, M.B., eds. Humana Press, Totowa, NJ, pp. 71–89, 1999.
18. Barch, M. J., Knutsen, T., and Spurbeck, J. L. *The AGT Cytogenetics Laboratory Manual, 3rd ed*. Lippincott-Raven, New York, 1997.
19. Lawce, H. How Colcemid works. *The Journal of the Association of Genetic Technologists*, **28(1):**5–9, 2002.
20. Rooney, D.E. *Human Cytogenetics: Constitutional Analysis, 3rd ed*. Oxford University Press, Oxford, 2001.
21. Pflueger, S. M. V. Cytogenetics of spontaneous abortion, in *The Principles of Clinical Cytogenetics* Gersen, S. L. and Keagle, M. B., eds. Humana Ps, Totowa, NJ, pp. 317–343, 1999.
22. Mascarello, J. T. Clinical correlates of cytogenetic anomalies, in *The Cytogenetic Symposia, 1994* Kaplan, B.J. and Dale, K.S., eds. The Association of Cytogenetic Technologists, Burbank, CA, pp. 12-1–12-9, 1994.
23. Burgoyne, P. S., Holland, K., and Stephens, R. Incidence of numerical chromosome anomalies in human pregnancy estimation from induced and spontaneous abortion data. *Hum. Reprod.* **6:**555–665, 1991.
24. Powell, C. M. Sex chromosomes and sex chromosome abnormalities, in *The Principles of Clinical Cytogenetics* Gersen, S. L. and Keagle, M. B., eds.. Humana Ps, Totowa, NJ, pp. 229–258, 1999.
25. Mathur, A., Stekol, L., Schatz, D., MacLaren, N. K., Scott, M. L., and Lippe, B. The parental origin of the single X chromosome in Turner syndrome: lack of correlation with parental age or clinical phenotype. *Am. J. Hum. Genet.* **48:**682–686, 1991.
26. Skuse, D. H., James, R. S., Bishop, D. V. M., et al. Evidence from Turner's syndrome of an imprinted X-linked locus affecting cognitive function. *Nature*. **387:**705–708, 1997.
27. Pai, G. S., Lewandowski, R. C., and Borgaonkar, D. S. *Handbook of Chromosomal Syndromes*. Wiley–Liss, Hoboken, NJ, 2003.
28. Wang, J.C. Autosomal aneuploidy, in *The Principles of Clinical Cytogenetics,* Gersen, S. L. and Keagle, M. B., eds. Humana Ps, Totowa, NJ, pp. 157–190, 1999.
29. Mitelman, F. Clustering of breakpoints to specific chromosomal regions in human neoplasia. A survey of 5,345 cases. *Hereditas*, **104:**113–119, 1982.
30. Heim, S. and Mitelman, F. *Cancer Cytogenetics, 2nd ed.* Wiley–Liss, New York, 1995.
31. Mitelman, F., Johansson, B., and Mertens, F. *Mitelman Database of chromosomal aberrations in cancer*. Available at http://cgap.nci.nih.gov/chromosomes/mitelman.
32. Mitelman, F. *Catalog of Chromosome Aberrations in Cancer, 5th ed.* Wiley–Liss, New York, 1994.

15 Fluorescence *In Situ* Hybridization

A Major Milestone in Luminous Cytogenetics

SUNEEL D. MUNDLE AND ROBERT J. KOSKA

1. INTRODUCTION

The 21st century is witnessing emergence of a new era in medicine called "Genomic Disease Management." Chromosomes, which house the essential components of this new faculty of medicine, have gained scientific attention since the late 19th century. However, it took nearly 100 yr before the understanding of these vital molecules could be taken to a patient's bedside. The association of chromosomal abnormalities with congenital malformation disorders and cancer has been suspected since the late 1950s *(1–4)*. In particular, demonstration by these studies of trisomy of one of the smallest chromosomes in Down's syndrome (chromosome 21, as shown later), monosomy X in Turner's syndrome, and the chromosomal analysis of leukemias laid foundation for the coming years. It was the development of a variety of staining methods assigning uniqueness to individual chromosomes that really marked the beginning of genomic analysis (for review, see ref. *5*). As shown in Table 1, these techniques are collectively known as chromosome banding. A realization that using certain stains like Giemsa and Quinacrine mustards, the AT-rich and GC-rich regions of chromosomes could be distinguished into a unique banding pattern for every chromosome, indeed, invited clinical interest into this fascinating field *(6–7)*. Using a metaphase cell preparation, it became possible to clearly identify chromosomal abnormalities in clinical specimens. The banding techniques underscored the fact that the association of chromosomal abnormalities with disease might not be random or an epiphenomenon. A demonstration of two consistent chromosomal translocations—t(9;22) related to Philadelphia chromosome in chronic myeloid leukemia (CML) and t(15;17) found in acute promyelocytic leukemia (APL)—provided direct testimony for a specificity of chromosomal anomalies in hematological malignancies *(4,8,9)*. Ever since, chromosomal banding became the gold standard for both clinical and basic science studies in cytogenetics.

Many authors view the progress of cytogenetics in three phases: (1) the Prebanding era (until 1970), (2) the banding era (1970–1980), and (3) the present-day constantly evolving molecular cytogenetics era (post-1980) *(10,11)*. The roots of this astoundingly progressive molecular phase in the last two decades actually could be seen as early as 1969, when the concept of *in situ* oligonucleotide hybridization (ISH) was introduced by several groups simultaneously *(12–14)*. Albeit, the use of radioisotopic labeling and the autoradiographic visualization method of ISH restricted its general application. Between 1986 and 1988, two major advances illuminated the field of cytogenetics: (1) the development of nonradioactive fluorescent oligonucleotide probe labeling technique and (2) the construction of human-chromosome-specific libraries *(15–18)*. During this period, the concept of "interphase cytogenetics" was born, which offered a tremendous technical ease in the ISH protocol and reduced the assay time considerably *(17)*. The fluorescence *in situ* hybridization (FISH) has since come a long way, metamorphosing from a concurrent detection of several chromosomal abnormalities to the identification of previously uncharacterized abnormalities with multiplex-FISH (M-FISH) or spectral karyotyping (SKY™), further to genomic screening with FISH-based metaphase-comparative genomic hybridization and most recently to the solid-phase genomic DNA arrays (for review, see refs. *11 and 19*). The present chapter describes the principles of ISH and FISH and their potential clinical applications and, finally, introduces the seeds of future developments in FISH technology.

2. *IN SITU* HYBRIDIZATION

In situ hybridization was originally described as a generic technique to detect nucleotide sequences (DNA and/or RNA) in tissues or cell preparations. However, over a period of time, especially after the introduction of FISH, the former has almost become synonymous with detection of messenger RNA (mRNA). As elegantly reviewed by McNicol and Farquharson *(20)*, ISH could be applied to a wide variety of biological specimens with diverse visualization systems, including autoradiography, enzymecolorimetry, fluorometry, and chemiluminescence.

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

Table 1
Chromosomal Banding

Giemsa (G) banding: Generically, Giemsa preparation consists of methylene blue, azure B, azure A, azure C and thionin. Essentially they are thiazine molecules with 4, 3, 2, 1, 0 methyl groups, respectively. In addition, there is eosin in the mixture. The staining mechanism involves an interaction between ionically charged thiazins and the phosphate moiety of the DNA sugar phosphate backbone. The staining enhances the banding pattern attained due to the heterochromatin and eu-chromatin packaging and the sequence of GC-rich vs AT-rich regions unique to individual chromosomes. This is called G-banding, which is considered to illumninate the AT-rich regions that replicate relatively late in the S-phase during cell cycle.

Reverse (R) banding: A treatment with high-molarity phosphate buffer at high temperature and lower pH and/or staining with acridine orange, produces bands complementary to those revealed by conventional G-banding. This banding pattern termed R-banding primarily demonstrates GC-rich regions that replicate early in the S-phase of the cell cycle. This reverse Giemsa banding method is used to confirm deletions detected by G-banding.

Centromeric (C) banding: The centromeric constitutive heterochromatin seems to have a cogent binding with nuclear matrix proteins, and as a result, depurinating treatments at low or high pH could enhance the Giemsa staining of the centromeric region, resulting in C-banding, and the rest of the chromosome looking pale.

Quinacrine (Q) banding: The Q-banding employs quinacrine with which the AT-rich region fluoresces brightly, whereas the GC-rich region quenches its fluorescence. Another dye that produces even brighter fluorescence with AT- rich regions is Hoechest 33258. In both cases, a direct interaction with DNA is responsible for banding and no interference with the DNA binding proteins is assumed.

Source: ref. *5.*

3. GENERAL PROCEDURES FOR ISH

3.1. SPECIMEN FIXATION With regard to *in situ* molecular detection techniques, the most basic influencing factor is specimen fixation, which serves to preserve morphology and retain the necessary target nucleotide sequences. The common fixatives in use are formalin, Carnoy's fixative (methanol : acetic acid), Bouin's solution (picric acid), and ethanol. Of these, buffered formalin seems to best preserve DNA and mRNA in a tissue section *(21)*. Formalin brings about target protein and nucleic acid fixation via crosslinking the protein amino groups and hydroxymethylation of nucleic acids *(22)*. The efficiency of formalin to preserve especially the low abundance RNA species in a tissue could be further enhanced by adjusting the pH of formalin to alkaline conditions *(23)*. Buoin's solution is generally not recommended for ISH/FISH applications. For cytological preparations on the other hand, simple air-drying or gentle fixation with methanol or Carnoy's fixative appear to work comparably *(24)*. The fact emphasized by different studies comparing a variety of fixation methods is that an immediate fixation of a specimen preserves the target with utmost efficacy.

3.2. SPECIMEN PRETREATMENT Most laboratory cell preparations such as cultured cell lines or blood lymphocytes usually do not have contaminating debris or cytoplasm from lysed cells or extracellular matrix (ECM) components and so forth and, hence might not need any pretreatment before hybridization with the probe. However, most types of target cell in a clinical specimen such as amniocytes, which have a layer of contaminating proteins, might need a pretreatment for optimal hybridization results. Also, the tissues fixed with crosslinking agents like formalin need pretreatment to unmask the target nucleotide sequences. Usually, the pretreatment protocol involves one or more of the following steps: (1) permeabalization with isothiocyanate solution or mild periodic acid solution, (2) mild acid hydrolysis with 0.05 *M* to 0.1 *M* HCl, and (3) proteolytic enzyme treatment primarily with pepsin, trypsin, or proteinase K *(20)*. Previous studies have noted that the protease treatment for 20–30 min offers definite signal enhancement. However, an, overextended protease treatment could compromise cellular morphology *(25)*. Additionally, a pretreatment with two to three short cycles (3–5 min each) of microwave heating in citrate buffer (pH 6.0) has been shown to enhance the hybridization outcome *(25,26)*.

Clearly, several options are available for pretreatment. The combination of different options would need to be optimized in each individual laboratory, as the initial specimen processing conditions vary greatly from place to place. When the specimen is pretreated, a mild postfixation with low-concentration paraformaldehyde is highly desirable to prevent the excessive loss of a target sequence and/or occasionally of the specimen itself *(20)*. With cell preparations, additional dehydration through graded alcohol is beneficial.

3.3. HYBRIDIZATION AND VISUALIZATION A variety of probe designs have been successfully used in ISH as reviewed earlier by McNicol *(20)*. Briefly, the double-stranded DNA probes are generated by cloning and amplification. These large double-stranded probes require denaturation to obtain a single-stranded DNA. On the other hand, a single-stranded complementary DNA (cDNA) probe could be generated directly by reverse transcribing mRNA with a specific primer followed by amplification employing polymerase chain reaction (PCR). PCR amplification of a DNA template using antisense primer provides another option. The advantage of the cDNA probe lies in its ability to directly hybridize without prior denaturation. An example of a PCR-generated single-stranded cDNA probe is shown in Fig. 1a. Alternatively, antisenseRNA or synthetic oligonucleotide sequences could also be used as probes. Some studies indicate that the antisenseRNA probes might offer higher sensitivity than the single-stranded cDNA probes in detecting low-abundance mRNA species *(27,28)*. However this proposition should be validated for individual situations.

Regardless of the probe type, they could be directly labeled using nucleotides containing radioisotope as originally described or conjugated with fluorescent chemical labels. Such labels could either be incorporated during the probe synthesis or 3′-end labeled with terminal transferase following their synthesis. The most common radioisotopes used are ^{32}P, ^{33}P, ^{3}H, and ^{35}S. Barring the safety concerns with their use, these labels do score merits for the ease in labeling and sensitivity of target detection. The use of nonisotopic fluorescent tags such as rhodamine, fluorescein, and so forth overcome the safety concerns regarding radiotisotopic labels and, in fact, according to some studies might even match comparably in sensitivity *(29)*. Alternatively, the probes could be labeled at the 5′ end with

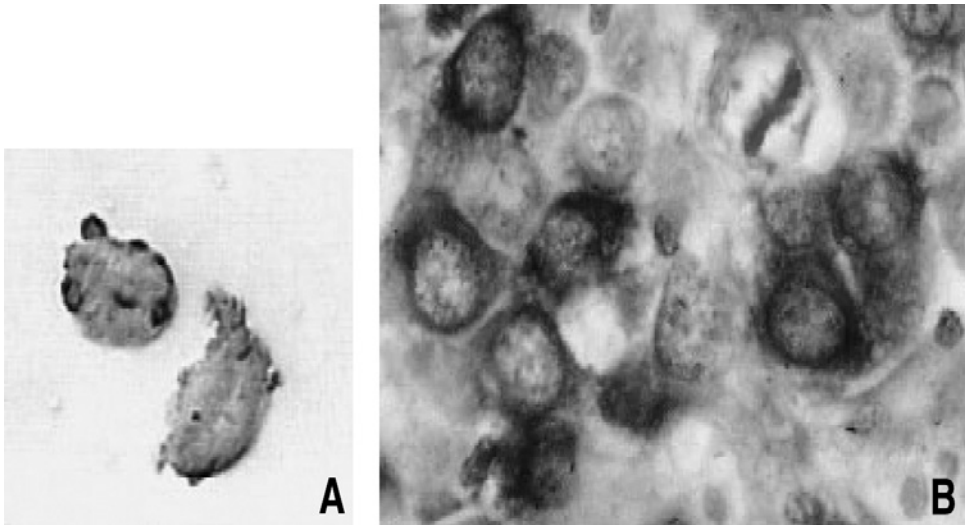


Fig. 1. *In situ* hybridization for detection of mRNA. **(A)** Cytomegalovirus (CMV)-infected human foreskin fibroblasts expressing mRNA (brown staining) for viral immediate early protein (IEP). A PCR-amplified single-stranded biotinylated cDNA probe prepared in the author's (SM) laboratory at Rush University, Chicago, IL, was employed with avidin–biotin–horse radish peroxide and diaminobenzidine as a detection system. Original magnification: x1000. **(B)** A paraffin-embedded breast cancer tissue section shows the presence of Histone-H3 mRNA (dark blue staining) (Courtesy; of Dr. Smitha Sivaraman, Rush University and Dr. Charuhas Deshpande and Dr. Sunil Badve of Northwestern University, Chicago) A synthetic commercial oligonucleotide probe tagged with fluoresceinisothiocyanate (FITC) was used, which was visualized by anti-FITC antibody conjugated to alkaline phosphatase and nitroso-blue substrate. Original magnification: ×1000. (Figure appears in color in insert following p. 172.)

protein haptens such as biotin and digoxigenin (indirect probe labeling). These reporter molecules could then be detected using fluorescent- or enzyme-labeled streptavidin in the case of biotin or a specific antibody in the case of digoxigenin. Figure 1a,b illustrates the use of ISH indirect reporting systems. The indirect probe labeling approach could at times compromise the sensitivity of target detection, and the use of blocking agents like bovine serum albumin (BSA) is required to block the nonspecific binding of haptens to cellular proteins *(30)*.

Generally, the factor that tremendously influences the specificity of probe binding is the hybridization condition of an assay *(31)*. The hybridization stringency is, in turn, determined by several factors like pH, salt and formamide concentration, temperature, and so forth. By varying annealing temperature and/or formamide concentration in the hybridization buffer, the nonspecific cross-hybridization could easily be prevented. Use of dextran sulfate is advocated for enhancing the efficiency of specific hybridization *(31)*. Usually, cRNA (antisenseRNA) probes appear to need longer hybridization times than cDNA probes. Competitive hybridization with an excess of unlabeled probe, probe preabsorption, use of sense-RNA probe, and RNAse or nuclease-treated specimens are examples of recommended controls for ISH *(20,31)*.

4. FLUORESCENCE *IN SITU* HYBRIDIZATION

Fluorescence *in situ* hybridization essentially indicates the fluorescent reporting of ISH. A sketch of routine FISH protocols for embedded tissue sections and cell preparations shown in Table 2 highlights its origin in the common ISH platform. FISH is commonly identified with its most popular application in cytogenetics, in contrast to the general notion of ISH as a RNA- detecting technique. In recent years, the growing commercialization of FISH technology has guided its way into the disease management armamentarium in a clinical laboratory. Although, the basic methodological considerations are carried over from ISH described earlier, several points of distinction deserve further description.

4.1. CELL CYCLE PHASE OF THE TARGET CELL As mentioned earlier, the major revolution in cytogenetics brought in by FISH was the possibility of cytogenetic analysis in any phase of the cell cycle, unlike the metaphase requirement of classical cytogenetic analysis by chromosomal banding. It was Cremer who coined the term "interphase cytogenetics" *(17,18)*. The metaphase analysis requires in vitro cell culture. Thus, it is time-consuming as well as labor-intensive and is often futile in certain tissue types. Surpassing this requirement in an interphase analysis not only offers a unique ease with time and effort but also makes it applicable to cytological and archived histological specimens equally well. The interphase analysis also allows for a simultaneous phenotype assessment. For instance, Rigolin et al. *(34)* showed that the monocyte-derived dendritic cells in patients with myelodysplastic syndromes carry the clonal cytogenetic abnormality. The same group later also showed a multilineage involvement in a 5q– syndrome using interphase FISH *(33)*. The metaphase analysis requiring dividing cells relies largely for accuracy on the proliferative potential of target cells, which varies significantly among different genotypic and phenotypic clones in a given clinical specimen. Thus, an inevitable preferential enrichment of rapidly proliferating clonal populations determine the eventual metaphase cytogenetics readings. This is obvious especially in the analysis of bone marrow disorders. The interphase analysis surmounting this difficulty could account for even minor cytogenetically abnormal clones *(36)*. The major limitation of interphase cytogenetics is the detection of only a specific abnormality and not permitting the localization of hitherto unidentified cytogenetic abnormality. Metaphase analysis scores

Table 2
General FISH Protocol

Steps involved in FISH	*Paraffin-embedded tissue sections*	*Cell preparations and isolated nuclei*
Intended application	Interphase cytogenetics	Interphase or metaphase cytogenetics
Fixatives	4–10% Formalin	Carnoy's fixative (methanol: acetic acid :: 3:1) or 4% formalin
Aging	None	Overnight or 2X SSC solution at 73°C for 2 min
Pretreatment	(a) Incubation at 52°C overnight (b) Deparaffinization using xylene or commercial agents like HemoDe and graded alcohol (c) Incubate sections in 1 *M* sodium thiocyanate (NaSCN) at 80°C for 20–30 min	None
Protease treatment	0.05% (wt/vol) pepsin in 0.01*N* HCl, pH 2.0, at 37°C for 10–20 min	0.05% (wt/vol) pepsin in 0.01 *N* HCl, pH 2.0, at 37°C for 10–20 min
Postfixation	4% Buffered formaldehyde	1% Buffered formaldehyde
Denaturation	Target: 70% formamide in 2X SSC, pH 7.0 at 72°C–74°C for 1–2 min Probe: 50–60% formamide in 2X SSC, pH 7.0 at 72°C–74°C for 1–2 min	Target: 70% formamide in 2X SSC, pH 7.0 at 72°C–74°C for 1–2 min Probe: 50–60% formamide in 2X SSC, pH 7.0 at 72°C–74°C for 1–2 min
Hybridization	Target incubation with probe prepared in a 2 X SSC solution containing 50–60% formamide and 10% dextran sulfate, pH 7.0 at 37°C–42°C overnight	Target incubation with probe prepared in a 2X SSC solution containing 50–60% formamide and 10% dextran sulfate, pH 7.0 at 37°C–42°C overnight
Posthybridization wash	(a)2X SSC with 0.3% NP-40, at 73°C for 2 min (b)2X SSC with 0.1% NP-40 at for 1 min	(a) 0.4X SSC with 0.3% NP-40, at 73°C for 2 min (b) 2X SSC with 0.1% NP-40 at for 1 min
Detection (necessary only for indirect FISH)	(1) Incubation with fluorescent labeled antibiotin or antidigoxigenin antibody (2) Alternative amplification with ezyme–Fluorophore– tyramide system	(1) Incubation with fluorescent labeled antibiotin or anti-digoxigenin antibody (2) Alternative amplification with ezyme–fluorophore– tyramide system
Counterstain/mounting	DAPI with antifade	DAPI with antifade
Visualization	Epifluorescence microscopy with appropriate filters	Epifluorescence microscopy with appropriate filters

Source: refs. *32* and *33.*

merit in this respect (for review, see ref.*32*). On the other hand, submicroscopic chromosomal rearrangements missed in banding analysis such as those known to occur in chronic myelogenous leukemia (CML) could be visualized by conducting FISH on metaphase and/or interphase cells *(37)*.

4.2. DNA PROBES Generally, most commercially available DNA FISH probes are prepared by selecting clones of yeast, bacterial, or P1 artificial chromosomes (YACs, BACs, or PACs respectively). Some probes are also prepared by PCR amplification of DNA. As shown in Table 3, there are primarily three types of chromosomal DNA probes that target the centromeric region, the entire length of chromosome, or the specific region of a particular chromosome *(38)*. The common application of the repetitive sequence centromeric probes for detecting numerical chromosomal abnormalities have earned them the name "chromosome enumeration probes" or CEP®s. The other two are named after their target range as the whole chromosome paints (WCP®) or locus-specific identifiers (LSI®), respectively. Additionally, owing to the importance of subtelomeric and telomeric regions in chromosomes in some cancers and congenital disorders, special LSI probes are available for these regions as well. Figure 2 shows photomicrographs illustrating different chromosome region targets of FISH probes detailed earlier. Further, the design of LSI probes targeting specific known balanced translocations has undergone considerable evolution in the last decade, minimizing the false-positive and/or false-negative rates. This is most apparent in the design of probes for t*(9;22)* translocation that results in a pathognomonic fusion *BCR-ABL* gene in CML *(39)*. The shift in the probe design for LSI BCR-ABL from independent fluorophore-labeled probes targeting only one side of the break point on the two chromosomes, resulting in a single fusion signal to the probes spanning both sides of the breakpoints on both chromosomes and resulting in two fusion signals has remarkably improved the specificity and considerably reduced the false-positive rate to <1% *(37)*. To assess the rearrangement of a gene with several translocation partners, the LSI design (break-apart probe) could include a combination of two probes labeled with distinct fluorophores and specific for sequences 5′ or 3′ of the breakpoint region within the rearranging gene. The separation of the two colors would thus be indicative of a rearrangement. Figure 2d,e shows examples of two unique LSI probe designs. When determining gene copy number, it is utmost important to have a built-in hybridization control. Usually, a centromeric or telomeric probe specific for the chromosome bearing the gene of interest is best suited as a built-in control. This is not only true for a gene deletion but is often needed in accurately determining the gene amplification status. In breast cancer, polysomy 17-associated increase in the HER2/*neu* gene copy number estimated using a combination of the HER2/*neu* LSI probe and a built-in CEP 17

Table 3
Comparison of Different Probe Designs

FISH probe	*CEP®*	*LSI®*	*WCP®*
Target	Pericentromeric alpha (and/or beta) DNA repeats	Specific DNA locus	Entire length of individual chromosome
Cell type	Metaphase or interphase	Metaphase or interphase	Metaphase only
Use of tissue sections	Possible	Possible	Not Possible
Major application	(a) Assessment of numerical chromosomal abnormalities	(a) Assessment of known chromosomal structural abnormalities	(a) Identification of structural abnormalities on a specific chromosome(s) of suspect
	(b) As built-in controls for LSI gene detecting deletion and amplification	(b) Detect locus specific abnormalities	(b) Identification of the origin(s) of a marker chromosome in a M-FISH assay
Chances of probe cross-hybridization	Possible under less stringent hybridization conditions	Nil	Negligible

control has been reported to relate in all likelihood to a normal protein expression pattern. Although a true amplification of the HER2/*neu* amplicon on chromosome 17, which is seen as a multifold increase over the chromosome copy number, results in abnormally high levels of HER2/*neu* protein and, in turn, has significant therapeutic implications.*(40,41)* Additionally, such built-in controls provide a parameter for determining nuclear truncation, especially in tissue sections, where only the nuclei with at least one of each signal type should be included in the final analysis.

In contrast to mRNA ISH described earlier, the cytogenetics FISH probes, which are mostly double stranded, require denaturation of probe and target prior to the hybridization step (Table 2). As emphasized by van Stedum and King *(32)*, the use of blocking unlabeled human DNA (e.g., Cot-1 DNA) and formamide in a probe mixture during denaturation/hybridization is essential to suppress the hybridization of probe to nonspecific repetitive DNA sequences within the target. Formamide is known to control the melting point of duplex nucleic acids *(32)*. The automated temperature-controlled protocols are available for denaturation/hybridization with equipments like HYBrite™.

4.3. FISH SIGNAL ENUMERATION In contrast to the ease in discerning the positivity of mRNA ISH, the signal patterns in cytogenetic FISH assays could be challenging and might need practice for perfection. Especially with LSI signals, the knowledge of probe design and the understanding of the specimen are extremely helpful. The factors that influence the quality and clarity of cytogenetic FISH signals are the adequacy of the amount of target DNA in a specimen, pretreatment of the specimen, denaturation/hybridization conditions, and duration, to name a few. An elaborate account of troubleshooting tips and signal enumeration guide can be found in the review by van Stedum and King *(32)*. A few salient points are as follows:

1. Excessive protease digestion or overdenaturation could damage morphology as well as cause extremely weak or even loss of signals. On the other hand, underdigestion with protease or other pretreatments could preserve morphology, but might compromise signal intensity.
2. Fluorescence *in situ* hybridization signals are very specific, relatively tiny, and might assume a round, oval, or diffuse shape depending on the chromosome target as well as the chromatin structure within that region.
3. A use of oil-immersion objective lens (×40, ×60, or ×100) and focusing up and down to view signals in all optical planes of the section are essential for accuracy.
4. Sometimes, a signal could appear split, which should be counted as one. In such cases, typically two very closely juxtaposed same fluorophore signals of approximately equal size are seen that are not separated by a distance larger than the diameter of either spot.
5. Overcrowded, superlayered, or inadequately counterstained areas of the specimen should be discounted from enumeration.
6. Use of the appropriate excitation/emission filters is a basic prerequisite.

Table 4 addresses a few points of common skepticism surrounding the FISH technique.

5. ADVANCEMENT OF FISH TECHNOLOGY

The constant quest for a genome-wide analysis has routed FISH technology through a remarkable progressive path. The major developments in this journey are described below.

5.1. MULTIPLEX/MULTICOLOR FISH OR SPECTRAL KARYOTYPING A simultaneous application of carefully chosen chromosome painting probes labeled with multiple fluorophores has made feasible to label all chromosomes in a single metaphase specimen. Both multicolor-FISH (M-FISH) and SKY serve the same purpose, but differ in their analysis procedure. In M-FISH, using selected narrow-bandpass filters separate images of each of the five fluorochromes are acquired and with a dedicated computational software that combines the unique labeling pattern of each chromosome, a distinct color index is determined for individual chromosome *(43)*. Alternatively, in SKY, a single image is obtained through a combination of a triple-bandpass filter and Fourier transform spectrometry. Finally, spectral analysis of individual pixels of the image provides the eventual 24-color display *(44)*. These two techniques leaped ahead of the classical metaphase banding cytogenetic analysis, in that they not only localize the marker chromosome but also trace the origins of the marker chromosome (Fig. 3). Additionally, M-FISH/SKY technology enables detection of cryptic chromosomal rearrangements as well. However, like classical cytogenetics, they suffer from the dependence on in vitro proliferation of target cells. A recent

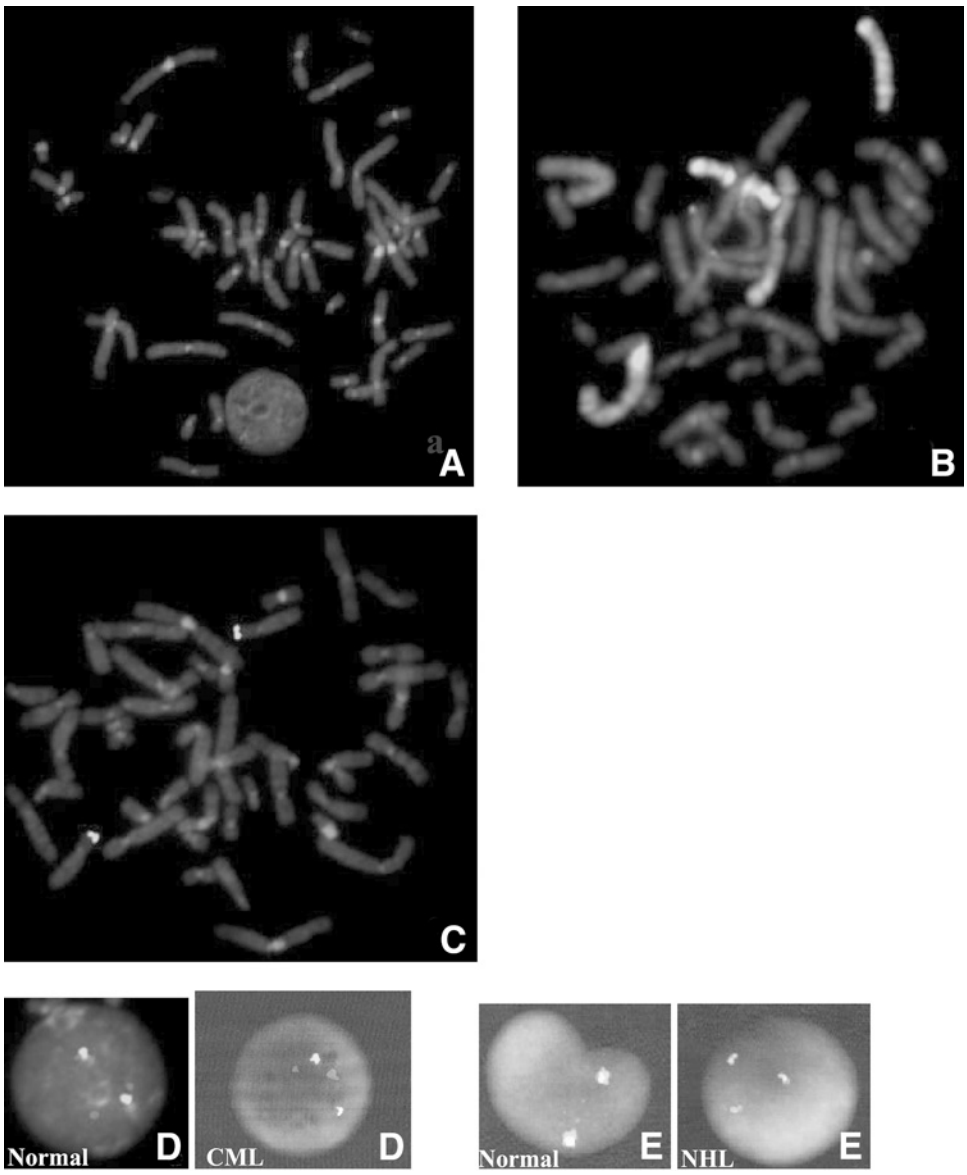


Fig. 2. Different types of DNA FISH probe. **(A)** CEP on normal metaphase and interphase- spectrum orange labeled centromeric probe for chromosome 11; **(B)** WCP on normal metaphase spread-simultaneous visualization of whole chromosomes 1, 2, and 4 in spectrum-green, spectrum-orange, and spectrum-aqua fluorophores, repectively; **(C)** subtelomeric probe for the p arm of chromosome 8 labeled with spectrum green. **(D)** LSI detection of bcr-abl using a design of a dual-color extra signal, which is a combination of spectrum-green LSI for bcr and spectrum-orange LSI for the abl-ASS region. Thus, a normal cell shows two orange and two green signals, whereas upon tanslocation of abl to bcr leaving behind intact ASS, in the bone marrow of a CML patient, one traslocation-specific fusion signal is seen along with a large orange signal for intact abl-ASS, a large green signal for intact bcr, and, in addition, a small extra orange signal for intact ASS left behind after bcr-abl translocation. **(E)** LSI break- apart design for detecting IgH rearragement. Because of the labeling of the probe in dual color (orange and green) in the same probe design, the normal nucleus in a non-Hodgkin's lymphoma bone marrow shows two fusion signals, whereas in an abnormal nucleus, the rearragement results in the separation of two fluorophores into one orange and one green signal along with the normal allele showing fusion signal. Original magnification: ×1000. (Parts a, b and c are the courtesy of the Quality Control department of Vysis, Inc., Downers Grove, IL; parts d and e are the courtesy of Dr. Wei-Tong Hsu, Rush Medical College, Rush University, Chicago, IL). (Figure appears in color in insert following p. 172.)

development in the principles of M-FISH/SKY is called combinatorial binary ratio labeling (COBRA-FISH). In this method, three fluorophores are used pair-wise to label a set of 12 chromosomes, each with a unique fluorochrome ratio. A fourth fluorophore is then added as a binary label to paint the second set of 12 chromosomes *(45).* Further addition of fluorophores has been proposed to provide extended painting options. Using COBRA-FISH, a recent report has shown new cryptic balanced translocations, t(7;17) (q32-34;q23) and t(7;17) (p15;q23) in accelerated phase/blastic crisis chronic myelogenonous leukemia (CML) *(46).*This approach, however, needs further investigation for sensitivity and specificity.

5.2. COMPARATIVE GENOMIC HYBRIDIZATION A genomewide screening of the gains or losses of the specific chromosomal regions came into sight with the development of comparative genomic hybridization (CGH). (for review, see refs. *38 and 47*). Importantly, such screening is possible from interphase cells, where CGH is superior over M-FISH, SKY, or

Table 4
Common Skepticism Regarding FISH

Q 1.	How stable are the probes? Can I freeze–thaw them? Can I store them diluted?
A:	Generally recommended storage temperature is –20°C, at which the probes are stable. They can be repeatedly frozen and thawed. Undiluted storage conditions are recommended for better preservation of the FISH probes
Q 2.	How long can I store my specimen slides and still hybridize them effectively?
A:	The slides stored at –20°C would hybridize with FISH probes, even after a period well over 6 mo.
Q 3.	Is sequential hybridization possible on the same specimen? Can FISH be combined with G-banding?
A:	Sequential hybridization with different probes is possible as well combination with G-banding possible. For technical details, see ref. *42*.
Q 4.	How long can the FISH signals be preserved following initial hybridization reaction?
A:	By storing slides at –20°C and minimizing the exposure to high-intensity ultraviolet light of the mercury lamp, the FISH signals could be preserved for months. Photobleaching is known to occur gradually over time and with frequent light exposure.
Q 5.	Do the CEP and WCP probes crossreact to other human chromosomes?
A:	Some crossreactivity with other human chromosomes may occur with the WCP probes. The crossreactivity could be reduced by using human Cot-1 DNA in the probe hybridization mixture.

COBRA-FISH. The technique of CGH relies on a competitive hybridization of test DNA and reference normal DNA that are differentially labeled with two fluorophores (typically Cy-3 with green fluorescence and Cy-5 with red fluorescence, respectively) and then hybridized concurrently in a 1 : 1 proportion to a target normal metaphase chromosome preparation. The computer-assisted quantification of the relative binding ratio of the test vs control over the entire length of each chromosome determines the loss or gain in a particular region in the test DNA. A major drawback of CGH is that if the size of the cytogenetically altered clone in a test specimen is small, it could easily be missed (Table 5). Therefore, in the case of solid tumors, to avoid normal cell contamination, a microdissection of a histopathologically well-defined tumor area is recommended. A plethora of reports are available on CGH analysis in solid tumors and hematological malignancies as reviewed by Zitzelsberger et al. *(47),* It is noted that a confirmatory test using FISH or PCR- based techniques is highly desirable to ascertain the abnormalities pointed by CGH.

A recent advance in CGH is seen in the development of CGH-based microarrays that use definitive DNA clones immobilized on a solid phase as the hybridization targets *(50,51)*. Although circumventing the reliance on metaphase and better defining the hybridization this technique entertains the possibility of high-throughput analysis. The recent genomic array contains approx 2400 target BAC clones to assess the DNA copy number across the human genome and has the ability to detect single-gene copy gains and losses in a homogeneous DNA specimen *(51)*.

5.3. FIBER-FISH The term "fiber-FISH" is used to describe high-resolution FISH-based genomic mapping of a less condensed, deproteinated, stretched chromatin fiber that is immobilized on a solid phase. A serial treatment of interphase cells on a glass slide with a cell lysis agent, deproteination in a high salt solution, followed by ultraviolet (UV) irradiation and air-drying gives long DNA fibers immobilized on the glass surface. Using FISH probes to hybridize with such DNA fibers, it is hoped that a fluorescent color bar code could be generated of about 400 kb in size that then could be used to screen the clinical DNA sample for gene rearrangements. The resolution of rearrangements of a few kilobases is thus expected to be achieved by fiber-FISH. *(52,53)*.

6. ISH/FISH SIGNAL AMPLIFICATION

In situ hybridization for low-abundance mRNA species as well as interphase cytogenetics signals are often very small and might need additional signal amplification for a meaningful visualization. There are two broad approaches for accomplishing ISH/FISH signal amplification (for a detailed review, see ref. *54*). In the first approach of 'target sequence *in situ* amplification,' the sequence of interest could itself be amplified *in situ* using a specific primer and labeled nucleotides, whereas in the other approach of posthybridization signal amplification, a single hybridization reaction is used with a reporter system that causes enhanced deposition of fluorochrome.

6.1. *IN SITU* AMPLIFICATION OF TARGET SEQUENCE

6.1.1. Primed In Situ Labeling The technique of primed *in situ* (PRINS) labeling described originally by Koch in 1992 *(55)* overcomes the occasional problems with cross-hybridization of CEP probes resulting from the homology of alphoid repetitive sequences among different chromosomes (e.g., chromosomes 13 and 21). A specific primer is used for the region of interest and it is elongated on 3′ end using labeled nucleotides and a thermostable *Taq* DNA polymerase. Repeated cycles of primed *in situ* synthesis over the target would enhance the signal in a process called "cycling PRINS." This method is considered to be rapid, and more sensitive than ISH or FISH. It is possible to apply PRINS to interphase nuclei and, thus, the possibility of using archival tissue materials demonstrate its clinical implications.

6.1.2. *In Situ* PCR Polymerase chain reaction amplification of a low copy number of viruses was first described by Haase in 1990 *(56)*. It is now possible to detect as low as one copy number of a nucleic acid sequence in a single cell using this technique of *in situ* PCR *(54)*. For PRINS and cycling PRINS or *in situ* PCR, again as seen with ISH/FISH, crosslinking fixatives like formalin are recommended because they retain the amplification product at the target sequence subcellular location, whereas the precipitating fixatives like Bouin's fluid tend to diffuse the product. Using specific primer(s) and *Taq* polymerase, an *in situ* amplification could be achieved either in a suspension of fixed cells in a regular thermocycler heating block or carried out on a special heating block in cells/tissues fixed on a glass slide. The detection of the amplified product then is carried out on a glass slide using a directly labeled FISH probe or a FISH signal reporting system if haptenized nucleotides were used in the PCR reaction. Despite initial enthusiasm, the technique at the end of the decade after its inception is still dwindling in use from routine clinical application because of the high false-positive and false-negative rates.

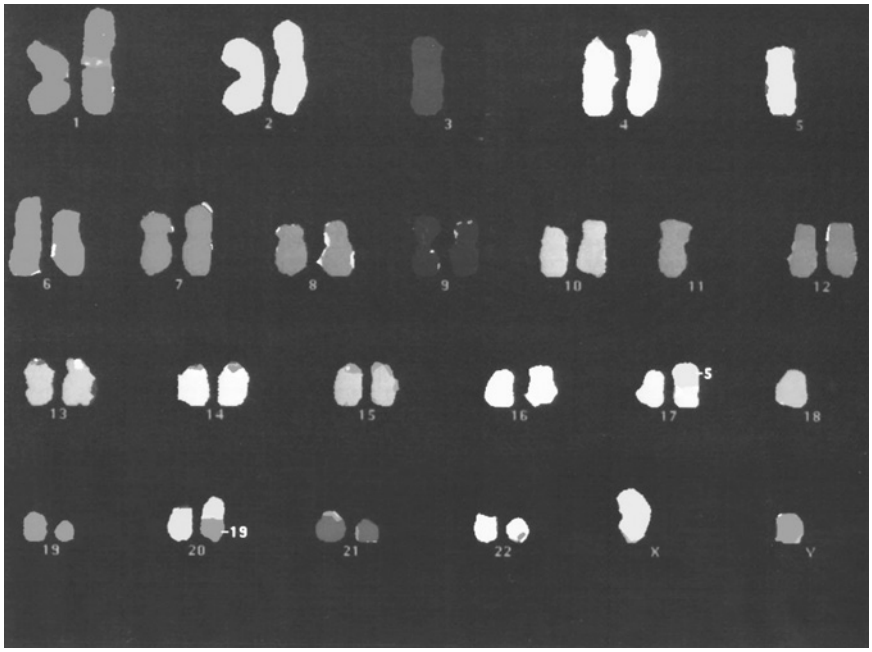


Fig. 3. M-FISH analysis of a metaphase obtained from the bone marrow of MDS patient. Note the complex karyotypic abnormalities with –3, –5,–11, and unexpected translocations t(5;17), and t(19;20) that could only be detected by the use of M-FISH. (Author's [SM] collaboration with Dr. Wei-Tong Hsu and Dr. Azra Raza at Rush University, Chicago, Illinois.) (Figure appears in color in insert following p. 172.)

Table 5
Comparison of the Overall Resolution of Different DNA Detection Techniques

Technique	*Resolution*
Metaphase chromosome banding	10 Mb
Metaphase FISH	1–10 Mb
M-FISH/SKY	~2–3 Mb
CGH	50 kb–10 Mb
Interphase FISH	50–100 kb
CGH-genomic matrix microarrays	40–100 kb
Fiber FISH	1–5 kb
Southern/Northern blotting	50–100 bp
PCR and nucleotide sequencing	1 bp

Source: refs. *48* and *49.*

At present, its major application is perceived to be in the detection of low-copy-number viruses in infected human cells.

6.2. POSTHYBRIDIZATION SIGNAL ENHANCEMENT

6.2.1. Tyramide Signal Amplification A novel catalyzed reporter deposition (CARD) system originally described by Bobrow et al. in 1989 *(57)* has undergone considerable evolution in the last decade. This system is considered to enhance the ISH signal 2.5- to 100-fold with diverse modifications described in the literature. In principle, this system uses horseradish peroxidase (HRP) either conjugated to antibody against the hapten label of the ISH probe as seen in the original reports or linked directly to the nucleotides as found in the impressive experiments of Raap and his co-workers in recent years *(58,59)*. The HRP catalyzes a covalent deposition of tyramide on tyrosine residues in the vicinity of enzyme action via a free-radical mechanism. The fluorescent labels of multiple tyramide residues thus deposited could be visualized by epifluorescence microscopy. An example of the tyramide amplification of ISH signals for low-abundance cytokine mRNA species is shown in Fig. 4. In addition to its application for detecting low-copy-number mRNA or viral DNA, the tyramide amplification system could also be used for small LSI probes to specifically detect small chromosomal translocations. Despite the major limitation of this system resulting from the as yet sole dependence on peroxidase catalysis, the attempts to use multiple probes applied sequentially and employing HRP through different means, a multicolor FISH chromosomal analysis has been made possible in recent reports *(59,60).*

7. CLINICAL IMPLICATIONS OF CYTOGENETIC ASSESSMENT BY FISH

Since the pioneering demonstrations of the association of chromosomal abnormalities with human diseases like Down's syndrome, Turner's syndrome, CML, APL, and so forth as described earlier, several other congenital disorders and cancers have been now linked with specific chromosomal abnormalities. With this knowledge and combined by the effective cytogenetic analysis tools, chromosomal banding and, among the sensitive molecular assays, FISH, genomic disease management is seeing the light of the day. The implications of cytogenetics in congenital disorders are primarily seen in diagnosis and in determining family history or, on a larger scale, in population genetics, whereas a significant advance is seen in cancer, where cytogenetics are used for diagnosis, prognosis, in selection of therapy, as well as for monitoring residual disease through and after therapy.

7.1. CORRELATION OF CYTOGENETICS WITH PATHOBIOLOGY OF THE DISEASE

The gene dosage effect causing several microdeletion syndromes has already been remarkably associated with definite phenotypic manifests. (for review, see ref. *61*). Fortunately, for many of these syndromes,

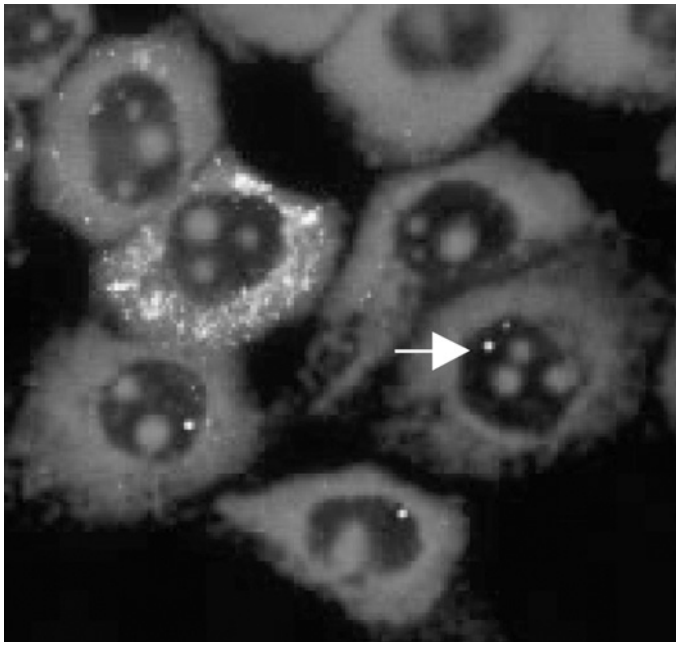


Fig. 4. Tyramide amplification of FISH signals. Interleukin-8 (IL-8) mRNA was detected in bladder carcinoma cell line 5637 using a HRP-conjugated specific oligonucleotide probe and a biotin–tyramide–antibody–fluorescein system as described previously *(59)*. On the background of the red counterstain resulting from the ISH for ribosomal RNA using Texas Red-labeled oligonucleotide probe, a clear cytoplasmic labeling (yellow) is seen localizing IL-8 mRNA. In addition, please note the green signals (arrow) for minor levels of IL-8 mRNA in the nucleus detected during synthesis *in situ*. Original magnification:x1000. (Courtesy of Dr. Anton K. Raap, Leiden University Medical Center, Leiden, The Netherlands). (Figure appears in color in insert following p. 172.)

commercial FISH tests are available to help in their definitive diagnosis. Similarly, certain chromosomal abnormalities have been found to be pathognomonic in certain hematological disorders. In general among all the cancers, a pathogenetic involvement of chromosomal abnormalities is well characterized in hematological malignancies compared to solid tumors. In cancer, both balanced and unbalanced aberrations have been described and it is proposed that these two types could be functionally very distinct, in that primary directly disease-related changes might be balanced and the secondary tumor progression-related changes might be unbalanced *(67)*. It appears that, today, knowledge of more than 600 neoplasia-related balanced chromosomal alterations exists, most of which, as mentioned earlier come from hematological malignancies *(63)*.

A direct link of a cytogenetic abnormality with specific tumor pathobiology has served to be pathognomonic in the diagnosis of some malignancies. As indicated by Mitelman *(63)*, indeed many of these typical aberrations result in a chimeric abnormal protein. The well-known examples of such chimeric proteins are constitutively active tyrosine kinase activities resulting from t(9;22) in CML(bcr-abl), or t(5;12) with the fusion of TEL-PDGFR in chronic myelomonocytic leukemia (CMML), or a transcription factor AML1/ETO resulting from t(8;21) in acute myeloid leukemia (AML) *(63,64)*. On the other hand, some aberrations lead to undesirable and untimely activation of a protein, such as bcl2, being turned on in follicular B-cell lymphoma as a result of t(14;18). Bcl-2, being antiapoptotic, causes a lymphoaccumulation disorder that eventually culminates into frank malignancy. Alternatively, sometimes it is the inactivation of protein that could result in increased survival and/or proliferation of cancer cells. In APL, t(15;17) inactivates the PML protein on chromosome 15, leading to a loss of its proapoptotic activity. Simultaneously, its fusion partner on chromosome 17, the retinoic acid receptor α (RARα) is rendered dysfunctional, causing a differentiation block in myeloid cells *(65)*. A direct pathological bearing of the cytogenetic aberrations in these disorders warrants their assessment by sensitive molecular techniques like FISH as a diagnostic aid. In addition, rapidity of FISH testing also would serve greatly for diagnostic purposes. A confirmation of the adherence of the abnormality to the expected chromosome, though, needs to be supported by the use of appropriate built-in chromosome localizing controls. The cytogenetic analysis would be invaluable especially when the biopsy histology is inconclusive such as the case reported for low-grade astrocytomas, which sometimes could pose a challenge for differential diagnosis from reactive gliosis. In this study, albeit anecdotal, chromosomal aneusomy was diagnostic in 30% inconclusive cases that were reported malignant only in a follow-up biopsy *(66)*.

7.2. RELATIONSHIP OF CHROMOSMAL ABNORMALITIES WITH DISEASE PROGNOSIS It is well established that the clinical manifests of a malignant disease vary tremendously in individual patients. Largely, such variability stems from a biological genotype of the malignant clone that determines the kinetics of disease progression and overall disease prognosis in an individual patient. Among the established biological clonal identifiers, cytogenetic abnormality is perhaps the most widely used marker besides the cell surface markers. Routine technique employed for this purpose is FISH. In myelodysplastic syndromes (MDSs), nearly 50% of the patients demonstrate single or complex cytogenetic abnormalities with frequent involvement of 5q– or –5, +8, –7, and 20q–. With the exception of 5q-, which shows low rates of leukemic transformation and better survival, all other abnormalities are known for poor survival. Also, it appears that an evolving complex karyotype corresponds with rapid leukemic transformation and poor survival *(67)*. Interestingly, although whole or partial chromosome losses or gains have been described for MDS, translocations have not been frequent, except for t(5;12) described recently in CMML, as mentioned earlier. However, often the metaphase analysis would note the presence of a marker chromosome. As shown in Fig. 3, the application of M-FISH certainly now helps in identifying the origin of such marker chromosome, further unfolding the biology of MDS. Because of such a cogent relationship of cytogenetic aberrations with leukemic trasformation and overall survival, cytogenetics are one of the three criteria considered in the recently proposed International Prognostic Scoring System (IPSS) for MDS risk classification *(68)*. In a recent report, similar differences in survival rates have been shown for distinct cytogenetic alterations in multiple myeloma in a large series of patients ($n = 351$) studied by the Eastern Cooperative Oncology Group (ECOG) of the United States *(69)*. In contrast to MDS, translocations involving 14q32 are very common in multiple myeloma. A remarkable

drop in survival rates were noted in patients with t(4;14)(p16;q32), t(14;16)(q32;q23), and –17p13. On the other hand, barring the –13q14 cases (intermediate group), all other cytogenetic abnormalities including t(11;14)(q13;q32) demonstrated nearly two fold higher survival rates and better prognosis. *(69)* Among solid tumors, a direct association has been shown between HER2/*neu* gene amplification on chromosome 17q and high rates of disease relapse and poor survival in node-positive metastatic breast cancer (for review, see ref. *70*)

7.3. CORRELATION OF KARYOTYPE WITH RESPONSE TO THERAPY HER2/*neu* amplification indeed seems to be a significant determinant in the response to different therapeutic regimens for breast cancer *(70).* Generally, in both node-negative and node–positive breast cancer, HER2/*neu* amplification relates to a resistance to preoperative chemotherapy with cyclophosphamide, methotrexate, and 5-fluorouracil (CMF). In contrast, the HER2/*neu*-amplified tumors could be more sensitive to the anthracyclin-based therapy regimens (doxorubicin + cyclophosphamide). Further, in a prolonged (20 yr) follow-up study on node-negative breast cancer cases treated with tamoxifen as adjuvant therapy in a randomized trial, HER2/*neu* amplification correlated with low estrogen receptor status, larger tumor size, and lower overall as well as disease-free survival *(71,72)*. The most significant implication of HER2/*neu* amplification has been demonstrated recently in the treatment of node-positive metastatic breast cancer with Herceptin®, a specific humanized antibody to HER2/*neu* protein receptor, as adjuvant chemotherapy. A remarkably higher overall response rate and disease-free survival was shown with Herceptin in combination with chemotherapy in cases of HER2/*neu*-amplified status *(73).*

Interestingly, in APL, the patients showing typical t(15;17) respond very well to all-*trans* retinoic acid (ATRA) differentiation therapy. However, the RARα at 17q12-21 in about 10% of APL cases shows variable partners (viz. promyelocytic leukemia zinc finger [*PLZF*] at 11q23, or nucleophosmin gene at 5q35, or nuclear mitotic apparatus gene at 11q13, or *STAT5b* on chromosome 17). It appears that the cases with t(11;17)(q23;q21) resulting in *PLZF*/RARα fusion do not respond to ATRA therapy as well unless combined with granulocyte-colony-stimulating factor (G-CSF) or hydroxyurea. The *PLZF*/RARα also do not seem to respond to arsenic trioxide (for a review of APL therapy, see ref. *65*). Clearly, cytogenetic analysis could guide the therapy selection at least for some cancers at this time.

7.4. CYTOGENETIC ANALYSIS IN DISEASE MONITORING Detection of a cytogenetically marked clone in a disease is of significant value in monitoring the effects of therapy, residual disease, and relapse. The major aim in such attempts is to determine biologic remission and to detect relapse in advance, thus bringing in a precision of much sought after tailor-made therapy. Both t(9;22) resulting in *bcr-abl* chimera in CML as well as t(15;17) causing PML-RARα fusion in APL have been successfully used to monitor effects of recently developed imatinib mesylate (Gleevec) therapy for CML and, as mentioned earlier, ATRA therapy for APL. As reviewed elegantly by Grimwade and Coco *(65)*, it was apparent from the UK Medical Research Council ATRA trial in APL, that persistence of PML-RARα at the end of the third course of therapy led to a poorer 5yr survival and showed almost twice as rapid rates of disease relapse in 5 yr. Furthermore, it has been shown that a pre-emptive therapy at the point of molecular relapse detected during postconsolidation therapy monitoring significantly improved a long-term survival than commencing therapy at hematological relapse, which was seen to occur within a range of 1–14 mo. The value of cytogenetic analysis in predicting the recurrence of disease has been seen in solid tumors as well. In a recent multicenter trial detecting aneusomy of chromosomes 3, 7, and 17 and loss of 9p21 in a single FISH assay in patients with prior diagnosis of transitional cell carcinoma of bladder, it was observed that finding cytogenetically abnormal transitional cells in voided urine corresponded to significantly high rates of recurrent disease. Interestingly, the predictive value of FISH in this respect was found to be superior to the most currently available other tests, including cytology *(74).*

The cytogenetic response criteria are most well defined for CML based on the proportion of reduction in the size of the abnormal clone with t(9;22). A *complete response* relates to undetectable abnormality, a *partial response* to 1–34% cells with t(9;22), *minor response* to 34–94% detection, and *no response* to a persistence of over 95% abnormal cells following therapy *(75).* A cytogenetic response to interferon-α treatment was shown to correlate with prolonged survival *(76)*, and FISH appears to be the assay of choice strongly recommended for monitoring such responses to CML therapies *(75).* In the phase I trial of a new therapeutic agent for CML, a tyrosine kinase inhibitor called STI 571, imatinib mesylate (Gleevec), 33% of the responding patients showed cytogenetic response evaluated by FISH *(77).* However, in young CML patients with bone marrow transplant as the front-line therapy, highly sensitive RT-PCR has been recommended for posttransplant disease monitoring despite some concerns of false-positivity *(75).*

8. FISH VS OTHER TECHNIQUES FOR CYTOGENETIC ANALYSIS

For the obvious reasons like high sensitivity, technical ease, applicability to a wide range of clinical specimens, and the feasibility of combining with phenotypic characterization, FISH has been welcomed in routine clinical diagnostic practices. FISH, while taking advantage of the foundation of cytopathological and histopathological competency, provides an additional spectacle revealing genomic secrets of the cellular pathology. This matrimony of cell and molecular biology is thus a major advantage offered by FISH.

The classical cytogenetic analysis with metaphase chromosomal banding continues to be a gold standard in genetic testing laboratories. However, several emerging studies report beyond the shadow of doubt that certain chromosomal abnormalities show a tendency not to grow well *in vitro* and/or result in poor spreading of chromosomes in metaphase preparations. Monosomy 7 in AML and MDS, t(15;17) in AML and +12 and 13q–/–13 in CLL were thus found to be frequently underestimated by conventional G-banding, and interphase-FISH provided the necessary sensitivity for their accurate estimation *(78).* The underestimation of t(15;17) and trisomy 12 cells by G-banding has been thought to result from their slow proliferative rates, leading to being outgrown in culture (*79,80*). On the other hand, a good correlation between G-banding and FISH

Table 6
A Comparative View of FISH, PCR, and RT-PCR

Characteristic	*FISH*	*PCR*	*RT-PCR*
Target	DNA	DNA	m-RNA
Detection principle	Complementarity hybridization	Complementarity and amplification	Complementarity, reverse transcription, and amplification
Sensitivity	1 in 10^3 cells	1 in 10^5 cells	1 in 10^5 cells
Specificity	Very high	Relatively high	Relatively high
Assay duration	1–2 d	1 d	1 d
False-positive rate	Negligible	Possibility because of amplification	Possibility because of amplification
False-negative rate	Possible if the abnormal clone size is extremely small	Negligible	Negligible
Chance cross-contamination	Highly unlikely	Likely	Likely
Quantitation	Possible (% abnormal cells)	Semiquantitative by densitometry or quantitative by real-time PCR	Semiquantitative by densitometry or quantitative by real time RT-PCR
Single-cell details and tissue architecture	Comprehensible	Unavailable	Unavailable
Application in general screening of genetic abnormalities	Possible in metaphase-FISH	Not possible	Not possible
Detection coverage of multiple chromosomal break-points	Easy	Cumbersome and requires multiple primer application	Cumbersome and requires multiple primer application
Detection of single-nucleotide changes	Not possible	Possible with additional nucleotide sequencing	Possible with additional nucleotide sequencing

Source: ref. *75.*

has been demonstrated in detecting t(9;22) in CML and t(8;21) in AML (*78*). Thus, rather than throwing the baby out with the bath water, such studies could guide the use of FISH along with chromosomal banding. This strategy is especially useful for evolving diseases like CML or MDS. While tracing the original cytogenetically marked clone by FISH, the metaphase analysis in these diseases could provide additional insights into evolving chromosomal aberrations that might be of prognostic significance. In this respect, an additional M-FISH analysis would further help in identifying the origin of the marker chromosome.

Polymerase chain reaction amplification technique extends the sensitivity of FISH further. However, as shown in Table 6, several crucial differences in the principles of these two assays warrant expert discretion in employing these assays judiciously. FISH seems to be limited in the detection level of 50–100 kbp change, whereas PCR combined with sequencing could potentially detect even a single basepair change *(48).* The major factors that affect the sensitivity of FISH are the probe size, labeling scheme, size and orientation of the target cell nuclei, and the number of nuclei enumerated for quantitation. In a recent study, it was found that the gap between FISH and PCR resulting from a relatively large size of the hybridization probe used and the lack of amplification could, at least in some instances, be bridged by using additional FISH on initial negative samples employing smaller probes *(81)*. In contrast, because PCR techniques rely heavily on the use of specific primers, often the assay could underestimate the regional rearrangement involving several breakpoints, which could easily be detected by FISH, as seen in a study detecting immunoglobulin light-chain mRNA in B-cell lymphoma *(82)*. Similarly, the scattering of the 11q13 breakpoints during t(11;14) in mantle cell lymphoma resulted in severely low detectability of the translocation with genomic PCR as compared to interphase-FISH *(83).* Another problem often noted with PCR/RT-PCR is a high frequency of false–positive cases, such as those reported by Gleißner et al. *(84)* in adult acute B-lineage lymphoblastic leukemia. FISH is thus warranted in such cases to rule out false-positivity. Furthermore, it might be noted that sometimes genetic changes can be detectable prior to their transcription into detectable levels of abnormal mRNA. It is especially relevant in posttherapy disease monitoring for predicting relapse. Quantitative real-time RT-PCR has been considered to be the assay of choice for monitoring the post-bone-marrow-transplant cases of CML *(75).* However, a study conducted in 2002 on a large serial specimen set from 78 posttransplant CML patients showed that, in some cases with relapse, the increase in FISH positivity preceded that of the quantititative RT-PCR *(85).* The genomic changes could also precede the appearance of detectable levels of protein as, often, FISH-positive cases are not matched by immunohistochemistry; for example, in mantle cell lymphoma, although t(11;14) was detected by FISH in 97% cases, only 69% were simultaneously positive for cyclin D1 protein immunoreactivity *(83).* As mentioned earlier, FISH also offers the luxury of following the biology of an abnormal clone through the malignancy evolution and posttherapy regression. Recently, a detection of telomerase mRNA by *in situ* PCR in cytogenetically abnormal clones localized simultaneously using various CEP probes demonstrated that a disappearance of telomerase-expressing abnormal clones in AML correlated with a response to chemotherapy *(86)*. Further, the feasibility of cell-by-cell analysis with FISH is of special value. PCR/RT-PCR techniques that use isolated nucleic acids could show false-negative results if they

have overpowering contamination from normal cells. The resolution to this problem is tissue microdissection, which at the moment is extremely tedious, demands special skills, and, hence, is less prevalent.

Clearly, FISH provides several unique advantages compared to the other molecular tools available. The choice of the molecular assay should rely on the clinical objective, ease of application, cost-effectiveness, nature of clinical specimen, and so forth. Also, the interpretation of data, especially in case of false-negative and false–positive cases, should be validated with another molecular technique. In the case of FISH, as also well demonstrated in cases of bcr-abl and PML-RARα, unbalanced deletions or insertion with or without duplication would result in variable labeling patterns that necessitate intelligible interpretation, as reported recently *(37,87)*.

9. FUTURE DIRECTIONS IN FISH-BASED ASSAYS

Currently, the major limitations of FISH is that it can only detect large numerical and structural changes and that it lacks the sensitivity of detection of a single nucleotide change as offered by PCR/sequencing. The development of so-called "padlock probes" offers the luxury of detecting single-basepair mismatch *(88)*. These are small oligonucelotide haptenized probes having extensions of oligonucleotides about 20 bp long at 3′ and 5′ ends, which are selected carefully to be juxtaposed upon their hybridization to the sequence of interest. The additional ligation reaction closes the ends of the padlock and attaches the probe covalently to the target. Because it requires two hybridization events over relatively very small stretches of targets, even a single-basepair mismatch hampers the hybridization of padlock probes. Perhaps their combination with amplification systems like fluorescent tyramides would improve their sensitivity further. Along with padlock probes, the amplification systems like PRINS, *in situ* PCR, and tyramide systems are expected to gain increasing prominence in FISH-based assays in the future. As mentioned earlier, clinical applications of FISH are already on the forefronts of hematology. However, what is exciting is the progress in solid tumor cytogenetics with mixtures of multiple probes and diverse sampling methods that could be expected to bring significant advances in solid tumor management (for review, see ref. *89*). A great promise is also felt in the utility of FISH- based assays for cancer risk screening, which will be ascertained in the next few years. Moreover, the technique of genomic microarray is in the wings to make an entry onto a routine clinical stage.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Smitha Sivaraman and Dr. Wei-Tong Hsu of the Rush University, Chicago, IL, and Dr. Scott Campbell, Dr. Irina Sokolova, Dr. Rolland Carlson, Ms. Wendy Bushen, Mr. Frank Fischinger, and Mr. Justin White of Vysis Inc., Downers Grove, IL for reviewing the manuscript and providing valuable critique. Thanks are also due to Ms. Gia Xenakis Branit, Vysis Inc., for her administrative support.

REFERENCES

1. Ford, C. E., Jacobs, P. A., and Lajtha, L. G. Human somatic chromosomes. *Nature (London)* **181:**1565–1568, 1958.
2. Ford, C. E., Jones, K. W., Polani, P. E., Almeida, J. C. D. E., and Briggs, J. H. A sex chromosomal anomaly in a case of gonadal dysgenesis (Turner's syndrome). *Lancet* **1:**711–713, 1959.
3. Lejeune, J., Gautier, M., and Turpin, R. Etudes des chromosomes somatiques de neuf enfants mongoliens. *C. R. Acad. Sci.* **248:** 1721–1722, 1959.
4. Nowell, P. C. and Hungerford, D. A. A minute chromosome in human chronic granulocytic leukemia. *Science* **132:**1497–1499, 1960.
5. Comings, D. E. Mechanisms of chromosome banding and implications for chromosome structure. *Annu. Rev. Genet.* **12:**25–46, 1978.
6. Caspersson, T., Lomakka, G., and Zech, L. The 24 fluorescence patterns of the human metaphase chromosomes-distinguishing characters and variability. *Hereditas* **67:**89–102, 1971.
7. Chaudhuri, J. P., Vogel, W., Voiculescu, I., and Wolf, U. A simplified method of demonstrating Giemsa-band pattern in human chromosomes. *Humangenetik* **14:**83–84, 1971.
8. Rowley, J. D. A new consistent chromosomal abnormality in chronic myelogenous leukemia. *Nature* **243**:290–293, 1973.
9. Rowley J. D., Golomb, H. M., Vardiman, J., Fukuhara, S., Dougherty, C., and Potter, D. Further evidence for a nonrandom chromosomal abnormality in acute promyelocytic leukemia. *Int. J. Cancer* **20:**869–872, 1977.
10. Sozzi, G., Testi, M. A., and Croce, C. M. Advances in cancer cytogenetics. *J. Cell. Biochem.* **32/33 (Suppl.)**, 173–182. 1999.
11. Kjeldsen, E. and Kølvaraa, S. FISH technique, FISH probes and their applications in medicine and biology-an overview, in *FISH Technology,* (Rautenstrauβ, B. and Liehr, T., edS.,) Springer-Verlag, Berlin, pp. 3–50, 2002.
12. Buongiorno-Nardelli, M. and Amaldi, F. Autoradiographic detection of molecular hybrids between rRNA and DNA in tissue sections. *Nature* **225:**946–947, 1969.
13. Gall, G. and Pardue, M. L. Formation and detection of RNA–DNA hybrid molecules cytological preparations. *Proc. Natl. Acad. Sci*., *USA* **63:**378–381, 1969.
14. John, H. L., Birnstiel, M. L., and Jones, K. W. RNA–DNA hybrids at the cytological level. *Nature* **223:** 912–913, 1969.
15. Pinkel, D., Straume, T., and Gray, J. W. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc. Natl. Acad. Sci, USA* **83:**2934–2938, 1986.
16. Deaven, L. L., Van Dilla, M. A., Bartholi, M. F., et al. Construction of human chromosome specific DNA libraries from flow sorted chromosomes. *Cold Spring Harb. Symp. Quant. Biol.* **51:**159–167, 1986.
17. Cremer, T., Landegent, J., Brückner, A., et al. Detection of chromosome aberrations in the human interphase nucleus by visualization of specific target DNAs with radioactive and non-radioactive in situ hybridization techniques: diagnosis of trisomy 18 with probe L1.84. *Hum. Genet.* **74:**346–352, 1986.
18. Pinkel D., Landegent, J., Collins, C., et al. Fluorescence *in situ* hybridization with human chromosome specific libraries: detection of trisomy 21 and translocation of chromosome 4. *Proc. Natl. Acad. Sci. USA* **85:**9138–9142, 1988.
19. Kearney, L. The impact of the new FISH technologies on the cytogenetics of hematological malignancies. *Br. J. Haematol.* **104:**648–658, 1999.
20. McNicol, A. M. and Farquharson, M. A. *In situ* hybridization and its diagnostic applications in pathology. *J. Pathol.* **182:**250–261, 1997.
21. Weiss, L. M. and Chen, Y. Y. Effects of different fixatives on detection of nucleic acids from paraffin-embedded tissues by *in situ* hybridization using oligonucleotide probes. *J. Histochem. Cytochem*. **39:**1237–1241, 1991.
22. Karlsen, F., Kalantari, M., Chitemerere, M. Johansson, B., and Hagmar, B. Modifications of human and viral deoxynucleic acid by formaldehyde fixation. *Lab. Invest.* **71:**604–611, 1994.
23. Basyuk, E., Bertrand, E., and Journot, L. Alkaline fixation drastically improves the signal of *in situ* hybridization. *Nucleic Acids Res.* **28:**e46, 2000.
24. Abati, A., Sanford, J. S., Fetsch, P., Marincola, F. M., and Wolman, S. R. Fluorescence *in situ* hybridization (FISH): a user's guide to

optimal preparation of cytologic specimens. *Diagn. Cytopathol*. **13:**486–492, 1995.
25. Oliver, K. R., Heavens, R. P., and Sirinathsinghji, D. J. S. Quantitative comparison of pretreatment regimens used to sensitize *in situ* hybridization using oligonucleotide probes on paraffin-embedded brain tissues. *J. Histochem. Cytochem*. **45:**1707–1713, 1997.
26. Henke, R. P. and Ayhan, N. Enhancement of hybridization efficiency in interphase cytogenetics on paraffin-embedded tissue sections by microwave treatment. *Anal. Cell Pathol.* **6:**319–325, 1994.
27. Qian, X., Bauer, R. A., Xu, H. S., and Lloyd, R.V. *In situ* hybridization detection of calcitonin mRNA in routinely fixed, paraffin-embdded tissue sections: a comparison of different types of probes combined with tyramide signal amplification. *Appl. Immunohistochem. Mol. Morphol.* **9:**61–69, 2001.
28. McQuaid, S., McMahon, J., and Allan, G. M. A comparison of digoxigenin and biotin labeled DNA and RNA probes for *in situ* hybridization. *Biotech. Histochem.* **70:**147–154, 1995.
29. Komminoth, P., Merk, F. B., Leav, I., Wolfe, H. J., and Roth, J. Comparison of ^{35}S- and digoxigenin-labeled RNA and oligonucleotide probes for *in situ* hybridization: expression of mRNA of the seminal vesicle protein II and androgen receptor genes in the rat prostate. *Histochemistry* **98:**217–228, 1992.
30. Brousset, P., Butet, V., Chittal, S., Selves, J., and Delsol, G. Comparison of *in situ* hybridization using different nonisotopic probes for detection of Epstein-Barr virus in nasopharyngeal carcinoma and immunohistochemical correlation with anti-latent membrane protein antibody. *Lab. Invest*. **67:**457–464, 1992.
31. Höfler, H. Principles of *in situ* hybridization, in *(In Situ) Hybridization: Principles and Practice,* Polak, J. M. and McGee, J.O'D., eds.) Oxford University Press, Oxford, pp. 15–30, 1990.
32. van Stedum, S. and King, W. Basic FISH techniques and troubleshooting. *Methods Mol. Biol.* **204:**51–63, 2002.
33. Werner, M., Wilkens, L., Aubele, M., Nolte, M., Zitzelsberger, H., and Komminoth, P. Interphase cytogenetics in pathology: principles, methods, and applications of fluorescence *in situ* hybridization (FISH). *Histochem. Cell Biol.* **108:**381–390, 1997.
34. Rigolin, G. M., Howard, J., Buggins, A., et al. Phenotypic and functional characteristics of monocyte-derived dendritic cells from patients with myelodysplastic syndromes. *Br. J. Haematol*. **107:**844–850, 1999.
35. Bigoni, R., Cuneo, A., Milani, R., et al. Multilineage involvement in the 5q- syndrome: a fluorescent *in situ* hybridization study on bone marrow smears. *Hematologica* **86:**375–381, 2001.
36. Wolfe, K. Q. and Herrington, C. S. Interphase cytogenetics and pathology: a tool for diagnosis and research. *J. Pathol*. **181:** 359–361, 1997.
37. Dewald, G. W. Interphase FISH studies of chronic myeloid leukemia. *Methods Mol. Biol*. **204:**311–342, 2002.
38. Tönnies, H. Modern molecular cytogenetics techniques in genetic diagnostics. *Trends Mol. Med*. **8:**246–250, 2002.
39. Lawce, H., Durum, C., Unsworth, N., Olson, S., and Magenis, R. E. BCR-ABL FISH: probes, patterns and prognoses. *J. Assoc. Genet. Tech.* **28:**40–46, 2002.
40. Wang, S., Saboorian, M. H., Frenkel, E. P., et al. Aneusomy 17 in breast cancer: its role in HER2/*neu* protein expression and implication for clinical assessment of HER2/*neu* status. *Mod. Pathol*. **15:**137–145, 2002.
41. McCormick S. R., Lillemoe, T. J., Beneke, J., Schrauth, J., and Reinartz, J. HER2 assessment by immunohistochemical analysis and fluorescence *in situ* hybridization. *Anat. Pathol*. **117:**935–943, 2002.
42. Jalal, S. M., Law, M. E., Christensen, E. R., Spurbeck, J. L., and Dewald, G. W. Method for sequential staining of GTL-banded metaphases with fluorescent-labeled chromosome specific paint probes. *Am. J. Med. Genet*. **46:**98–103, 1993.
43. Speicher, M. R., Ballard, S. G., and Ward, D.C. Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nature Genet.* **12:**368–375, 1996.
44. Schröck, E., du Manior, S. Veldman, T., et al. Multicolor spectral karyotyping of human chromosomes. *Science* **273:**494–497, 1996.
45. Tanke, H.J., Wiegant, J., van Gijlswijk, R.P., et al. New strategy for multi-colour fluorescence *in situ* hybridization: COBRA: combined binary ratio labeling. *Eur. J. Hum. Genet.* **7:**2–11, 1999.
46. Barbouti, A., Johansson, B., Hoglund, M., et al. Multicolor COBRA-FISH analysis of chronic myeloid leukemia reveals novel cryptic balanced translocations during disease progression. *Genes Chromosomes Cancer* **35:**127–137, 2002.
47. Zitzelsberger, H., Lehman, L., Werner, M., and Bauchinger, M. Comparative genomic hybridization for the analysis of chromosomal imbalances in solid tumors and haematological malignancies. *Histochem. Cell Biol.* **108:**403–417, 1997.
48. Caburet, S., Conti, C., and Bensimon, A. Combing the genome for genomic instability. *Trends Biotech.* **20:**344–350, 2002.
49. Saracoglu, K., Brown, J., Kearney, L., et al. New concepts to improve resolution and sensitivity of molecular cytogenetics diagnostics by multicolor fluorescence *in situ* hybrodization. *Cytometry* **44:**7–15, 2001.
50. Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., et al. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* **20:**399–407, 1997.
51. Snijders A. M., Nowak, N., Segraves, R., et al. Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat. Genet.* **29:**263–264, 2001.
52. Rapp, A.K., Florijn, R.J., Blonden, L.A.J., et al. Fiber FISH as a DNA mapping tool. *Methods* **9:**67–73, 1996.
53. Heng, H. H. Q. and Tsui, L. C. High resolution free chromatin/DNA fiber fluorescent *in situ* hybridization. *J. Chromatogr.* **806:**219–229, 1998.
54. Komminoth, P. and Werner, M. Target and signal amplification: approaches to increase the sensitivity of *in situ* hybridization. *Histochem. Cell Biol.* **108:**325–333, 1997.
55. Koch, J., Mogensen, J., Pedersen, S., et al. Fast one-step procedure for the detection of nucleic acids *in situ* by primer-induced sequence-specific labeling with fluorescein-12-dUTP. *Cytogenet. Cell Genet*. **60:**1–3, 1992.
56. Haase, A. T., Retzel, E. F., and Staskus, K. A. Amplification and detection of lentiviral DNA inside cells. *Proc. Natl. Acad. Sci., USA* **87:**4971–4975, 1990.
57. Bobrow, M. N., Harris, T. D., Shaughnessy, K. J., and Litt, G. J. Catalyzed reporter deposition, a novel method of signal amplification. Application to immunoassay. *J. Immunol. Methods* **125:**279–285, 1989.
58. van Gijlswijk, R. P. M., Zijlmans, H. J. M. A. A., Wiegant, J., et al. Fluorochrome-labeled tyramides: use in immunohistochemitry and fluorescence *in situ* hybridization. *J. Histochem. Cytochem*. **45:**375–382, 1997.
59. van de Corput, M. P. C., Dirks, R. W., et al. Sensitive mRNA detction by fluorescence *in situ* hybridization using horseradish peroxidase-labeled oligodeoxynucelotides and tyramide signal amplification. *J. Histochem. Cytochem*. **46:**1249–1259, 1998.
60. Speel, E. J. M., Ramaekers, F. C. S., and Hopman, A. H. N. Sensitive multicolor fluorescence *in situ* hybridization using catalyzed reporter deposition (CARD) amplification. *J. Histochem. Cytochem*. **45:**1439–1446, 1997.
61. Stumm, M., Tönnies, H., and Wieacker, P. F. Molecular cytogenetics techniques for the diagnosis of chromosomal abnormalities in childhood disease. *Eur. J. Pediatr.* **158:**531–536, 1999.
62. Johansson, B., Mertens, F., and Mitelman, F. Primary vs. secondary neoplasia-associated chromosomal abnormalities—balanced rearrangements vs. genomic imbalances? *Genes Chromosomes Cancer* **16:**155–163, 1996.
63. Mitelman, F. Recurrent chromosome aberrations in cancer. *Mutat. Res.* **462:**247–253, 2000.
64. Slupianek, A., Hoser, G., Majsterek, I., et al. Fusion tyrosine kianses induce drug resistance by stimulation of homology-dependent recombination repair, prolongation of G2/M phase, and protection from apoptosis. *Mol. Cell. Biol.* **22:**4189–4201, 2002.
65. Grimwade, D. and Coco, F. L. Acute promyelocytic leukemia: a model for the role of molecular diagnosis and residual disease monitoring in directing treatment approach in acute myeloid leukemia. *Leukemia* **16:**1959–1973, 2002.

66. Wessels, P. H., Hopman, A. H. N., Ummelen, M. I .J., Krijne-Kubat, B., Ramaekers, F. C. S. and Twijnstra, A. Differentiation between reactive gliosis and diffuse astrocytoma by *in situ* hybridization. *Neurology* **56:**1224–1227, 2001.
67. Padua R. A., McGlynn, A., and McGlynn, H. Molecular, cytogenetics and genetic abnormalities in MDS and secondary AML, in *Myelodysplastic Syndromes and Secondary Acute Myelogenous Leukemia*, Raza, A. and Mundle, S., eds., Kluwer, Boston, pp. 111–158, 2001.
68. Greenberg, P. L., Cox, C., LeBeau, M. M., et al. International scoring system for evaluating prognosis in myelodysplastic syndromes. *Blood* **89:**2079–2088, 1997.
69. Fonseca, R., Blood, E., Rue, M., et al. Clinical and biological implications of recurrent genomic aberrations in myeloma. *Blood*. **101:**4569–4575, 2003.
70. Masood, S. and Bui, M. M. Prognostic and predictive value of HER2/*neu* oncogene in breast cancer. *Microsc. Res. Tech*. **59:**102–108, 2002.
71. Carlomagno, C., Perrone, F., Gallo, C., et al. c-erb B2 overexpression decreases the benefit of adjuvant tamoxifen in early-stage breast cancer without axillary lymph node metastases. *J.Clin. Oncol*. **14:**2702–2708, 1996.
72. Bianco, A.R., De Laurentiis, M., Carlomagno, C.,et al. 20 year update of the Naples GUN trial of adjuvant breast cancer. *Proc. Am. Soc. Clin. Oncol*. **17:**373, 1998.
73. Slamon, D., Leyland-Jones, B., Shak, S., et al. Addition of Herceptin (humanized anti-HER-2 antibody) to first line chemotherapy for HER2 overexpressing metastatic breast cancer markedly increases anticancer activity: A randomized multinational controlled Phase III trial. *Proc. Am. Soc. Clin. Oncol*. **17:**377, 1998.
74. Sarosdy, M. F., Schellhammer, P., Bokinsky, G.,et al. Clinical evaluation of a multi-target fluorescent *in situ* hybridization assay for detection of bladder cancer. *J. Urol.* **168:**1950–1954, 2002.
75. Wang, Y. L., Bagg, A., Pear, W., Nowell, P. C., and Hess, J. L. Chronic myelogenous leukemia: laboratory diagnosis and monitoring. *Genes Chromosomes Cancer* **32:**97–111, 2001.
76. Kantarjian, H. M., Smith, T. L., O'Brien, S., Beran, M., Pierce, S., and Talpaz, M. Prolonged survival in chronic myelogenous leukemia after cytogenetics response to interferon-alpha therapy. The leukemia service. *Ann. Intern. Med.* **122:**254–261, 1995.
77. Druker, B. J., Talpaz, M., Resta, D., et al. Clinical efficacy and safety of an ABL specific tyrosine kinase inhibitor as targeted therapy for chronic myelogenous leukemia. *Blood* **94(Suppl.)**, 368a, 1999.
78. Tanaka, K., Arif, M., Eguchi, M., et al. Interphase fluorescence *in situ* hybridization overcomes pitfalls of G-banding analysis with special reference to underestimation of chromosomal aberration rates. *Cancer Gene. Cytogenet*. **115:**32–38, 1999.
79. Manicini, M., Nani, M., Cedrone, M.,et al. Combined cytogenetics, FISH and molecular analysis in acute promyelocytic leukemia at diagnosis and complete remission. *Br. J. Haematol.* **91:**878–884, 1995.
80. Knuutila, S., Elonen, E., Teerenhovi, L., et al.Trisomy 12 in B cells of patients of B-cell chronic lymphocytic leukemia. *N. Engl. J. Med.* **314:**865–869, 1996.
81. Grimwade, I. S., Chase, D., Goldstone, A., Burnett, A. H., Goldman, A. K., and Swirsky, J. M. Identification of PML/RAR alpha rearrangements in suspected acute promyelocytic leukemia (APL) using *in situ* hybridization on bone amrrow smears: a comparison of cytogenetics with RT-PCR in MRC ATRA trial patients. *Br. J. Haematol.* **101(suppl.):**35, 1998.
82. McNicol, A. M., Farquharson, M. A., Lee, F. D., and Foulis, A. K. Comparison of *in situ* hybridization and polymerase chain reaction in diagnosis of B cell lymphoma. *J. Clin. Pathol.* **51:**229–233, 1998.
83. Belaud-Rotureau, M. A., Parrens, M., Dubus, P., Garroste, J. C., de Mascarel, A., and Merlino, J. P. A comparative analysis of FISH, RT-PCR, PCR and immunohistochemistry for the diagnosis of mantle cell lymphomas. *Mod. Pathol.* **15:**517–525, 2002.
84. Gleißner, B., Rieder, H., Thiel, E., et al. Prospective BCR-ABL, analysis by polymerase chain reaction (RT-PCR) in adult acute B-lineage lymphoblastic leulemia: reliability of RT-nested PCR and comparison to cytogenetics data. *Leukemia* **15:**1834–1840, 2001.
85. Kim, Y. J., Kim, D. W., Lee, S., et al. Comprehensive comparison of FISH, RT-PCR, and RQ-PCR for monitoring the BCR-ABL gene after hematopoietic stem cell transplantation in CML. *Eur. J. Haematol*. **68:**272–280, 2002.
86. Gladstone, B., Sivaraman, S., Venugopal, P., et al. Measurement of telomerase activity in cytogenetically marked AML cells could be useful in determining response to therapy. *Blood* **98:**2439, 2001.
87. Ambre-Kadam, P. S., Baisane, C., Saikia, T., Nair, R., Gawade, H., and Advani, S. Fluorescence *in situ* hybridization: a highly efficient technique of molecular diagnosis and prediction for disease course in patients with myeloid leukemias. *Cancer Genet. Cytogenet.* **131:**125–134, 2001.
88. Nilsson M., Krjeci, K., Koch, J., Kwaitkowski, M., Gustavson, P. and Landegren, U. Padlock probes reveal single-nucleotide differences, parent of origin and in situ distribution of centromeric sequences in human chromosomes 13 and 21 *Nat. Genet*. **16:**252–255, 1997
89. Varella-Garcia, M. Molecular cytogenetics in solid tumors: laboratorial tool for diagnosis, prognosis and therapy. *Oncologist* **8:**45–58, 2003.

16 Immunohistochemistry

JOHN HUNT, LARISSA DAVYDOVA, RICHARD W. CARTUN, AND MARIA BAIULESCU

1. INTRODUCTION

Since its introduction as a routine diagnostic procedure over 25 yrs ago, immunohistochemistry (IHC) has revolutionized the field of surgical pathology. This powerful technique allows greater precision in the characterization and diagnosis of solid tumors, hematolymphoid neoplasms, and infections than ever before. An increasing number of antibodies directed against normal and abnormal cellular proteins as well as infectious agents is available to the surgical pathologist to diagnose and subclassify disease entities. These markers can be used in a variety of diagnostic and research settings.

We live in a time when a number of diseases can be characterized by a single genetic alteration that is easily assayed in the modern molecular pathology laboratory. Unfortunately, a laboratory with this level of sophistication is not yet readily accessible to the majority of practicing pathologists. Likewise, there are few pathologists that are trained in the performance and evaluation of molecular studies. For surgical pathologists, it is IHC, a test that focuses on recognition of protein products expressed by different cell populations in conjunction with a morphologic examination, that is used as a means to circumvent the need for direct evaluation of nucleic acid alterations. With this method, the products of genes are assayed in tissue sections, often allowing one to characterize a cell population as benign or neoplastic, determine cell lineage, and, in some cases, even determine the nature of the molecular genetic alteration leading to the process.

As with so many areas of medicine, when something seems too good to be true, it probably is. Immunohistochemistry is not a panacea that can provide a concise answer in every questionable case. In fact, if its use is not guided by a working knowledge of morphology, immunohistochemical methods, and their limitations, IHC can be as much a hindrance as a help. In this chapter, we will discuss the history and methodology of IHC, provide a brief introduction to the evaluation of tumors with immunohistochemical studies, and review some specific examples of how it correlates with molecular genetics.

2. HISTORY

Immunohistochemistry has been around in its rudimentary forms since the 1940s, when Coons experimented with immunofluorescence to detect antigens in frozen tissue sections *(1)*. Unfortunately, immunofluorescence usually requires fresh frozen tissue and lacks the sensitivity required for routine surgical pathology applications. The use of immunohistochemical methods on routinely processed surgical tissue sections was first reported by Taylor and colleagues in the mid-1970s *(2)*. The introduction of IHC into daily practice has taken some time, but most anatomic pathologists today would be loath to practice without access to at least some immunohistochemical markers.

A large step forward in the development of IHC techniques occurred in 1975 with the development of hybridoma cell lines capable of producing large quantities of pure monoclonal antibodies *(3)*. Also in 1975, the use of enzyme digestion to help unmask antigens altered by formalin fixation was described by Huang et al. *(4)*. These humble beginnings led to an explosion of literature and a rapidly growing list of IHC "stains" that could delineate cell lineages, specific products of cells, and, in some cases, the difference between benign and malignant cells. With experience and constant investigation, antibodies with greater specificity have been developed and newer techniques such as heat-induced epitope retrieval (HIER) have improved the sensitivity *(5)*. Other advances that improve laboratory logistics such as automated staining platforms, commercial antibody "kits," and polymer detection techniques have improved standardization, decreased turnaround time, and reduced nonspecific staining. A dizzying array of options is currently available for routine diagnostic use in many pathology laboratories.

3. METHODOLOGY

A complete discussion of antibody production and IHC methodology is beyond the scope of this chapter. For further information, the reader is referred to recent textbooks of immunology and IHC *(6,7)*.

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

3.1. ANTIBODIES Antibodies are proteins produced by B-lymphocytes in response to an antigen, which is a part of a protein, carbohydrate moiety, or some other biochemical construct that can be recognized and bound by an antibody. An immunogen, by contrast, is something that, when encountered by the immune system, stimulates the production of an antibody. Antibodies are immunoglobulin molecules that comprise two identical heavy and two identical light chains, each of which includes a constant region and a variable region. The amino acid sequence of the variable region confers steric and electrochemical properties that help determine how tightly (avidity) and how specifically (affinity) an antibody will interact with an antigen.

Antibodies for immunohistochemical studies are produced by animal hosts in response to an antigen stimulus. Depending on the method of production, antibodies can be monoclonal (produced in a hybridoma cell line in an animal host) or polyclonal (produced by many different cells in a live animal). As the names imply, the monoclonal antibody is derived from a single clone and produces a single antibody that recognizes one epitope of an antigen, whereas polyclonal antibodies are made by many cells that may recognize different epitopes on the antigen used to stimulate production. Each form has inherent advantages and disadvantages. The monoclonal antibody, usually produced in a mouse host, has the advantages of homogeneity, reproducibility, and lack of admixed nonspecific antibodies. Polyclonal antibodies will have lot-to-lot variability and usually higher background staining because they are not made by a single clone and they recognize multiple epitopes on a given antigen. The latter property can be advantageous if the antigen of interest is susceptible to degradation or alteration during fixation and also because the conditions under which the stains are used are less stringent. Although monoclonal antibodies are frequently preferred because of to specificity and reproducibility, several caveats must be considered. For instance, some antibodies are raised against fresh cells that have antigens that degenerate or are masked as a result of to suboptimal processing conditions or of fixation.

In IHC, antibodies serve two purposes. The first function is as a probe to recognize and bind a specific epitope such as the intermediate filament, cytokeratin, or the calcium-binding protein, S-100, and so forth that is mediated through the variable region. This function is characteristic of the "primary" antibody. Antibodies also serve as antigens wherein the constant region of the primary antibody is recognized by a "secondary"antibody conjugated to a fluorescent dye, or to biotin, which, in turn, binds an avidin–enzyme complex, or directly to an enzyme that can produce a localized visible result when reacted with a chromogenic substrate. Allowing visualization of the anatomic and histologic distribution of the antigen of interest is one of the main advantages of this technique.

3.2. IMMUNOHISTOCHEMICAL TECHNIQUES It is not only the specificity of the primary antibody that determines the usefulness of an antibody. Other factors related to the tissue and its processing are equally important. These include the fixative, duration of fixation, tissue processing, the method of antigen retrieval, and the detection system, to name a few *(8)*. The last critical ingredient is a working knowledge of morphology and expected reactions so that results can be interpreted as normal, abnormal, or indeterminate.

Attempts to find alternative fixatives to formalin for preservation of tissue morphology and immunoreactivity have been largely unsuccessful. Formalin, which is inexpensive, is easy to use, and provides good and reproducible morphology, will likely remain the standard fixative used in surgical pathology for the foreseeable future. The fact that current morphologic criteria for diagnoses are based on the appearance of tissues fixed in formalin, embedded in paraffin, and stained with hematoxylin and eosin makes the possibility of developing and using other fixatives even less likely. At the molecular level, formalin fixation results in the formation of crosslinks between proteins or between proteins and nucleic acids, involving hydroxymethylene bridges *(9)*. As a result of the alteration of the three-dimensional structure of the proteins, some antigenic sites are masked and become inaccessible to antibodies used in IHC. Initial efforts to find a better fixative than formalin were redirected after the discovery that at least some of the changes induced by formalin were reversible with the use of antigen retrieval techniques *(10)*. It is interesting to note that these same techniques might be able to improve yield in nucleic acid-based tests as well *(11)*.

3.2.1. Antigen Retrieval Antigen retrieval (AR) can be a simple and effective method that allows antigens masked during formalin fixation to be recognized by antibodies, but extensive testing is necessary to optimize the retrieval for different types of antibody. The first method for unmasking antigens in formalin-fixed tissue involved the use of proteolytic enzymes *(12)*. A number of enzymes have been used, including trypsin, pronase, pepsin, and ficin. However, the number of variables that have to be considered for enzymatic epitope retrieval make this technique less desirable. Different antibodies require different enzyme techniques, and inappropriate use of proteases could result in false-negative reactions and tissue destruction. The choice of enzyme also depends on the type of fixative used. Furthermore, the length of enzymatic digestion should be proportional to the extent of exposure to fixative, which is difficult to assess and control. Although HIER has taken a leading role in antigen retrieval, enzyme retrieval techniques are still preferred for detection of some antigens *(13)*. The requirements for pretreatment must be established by the institution preparing the immunohistochemical stains, as conditions are highly variable among different laboratories.

The antigen retrieval technique introduced by Shi et al. in 1991 *(5)* has revolutionized IHC and greatly expanded the number of antigens that can be detected in formalin-fixed, paraffin-embedded tissue *(10,14,15)*. In HIER, the unmasking of antigens is brought about by placing tissue in different types of buffer at temperatures between 80°C and 125°C. Studies have shown that the product of temperature and time, and the pH of the buffer used (rather then its chemical composition) are the major factors that influence antigen retrieval *(8)*. Different heating methods (microwave, pressure cooker, water bath, steamer, or autoclave) can be adjusted to yield similar intensities of staining. Water baths that use a lower temperature (90–95°C) require a longer time but show better preservation of morphology, which is critical to interpretation of results.

The most widely used retrieval buffers are citrate-based at pH = 6.0. The use of high-pH buffers in conjunction with high temperature for low-expression tissue antigens or for overfixed tissue enhances the sensitivity of detection, at the risk of increased tissue loss (tissue tends to detach from the slide with this method) and increased background staining.

The introduction of HIER has increased the sensitivity of staining for a wide range of antibodies and has made possible the retrieval of some antigens that are otherwise negative even with other unmasking pretreatments (Ki-67 [clone MIB-1], bcl-2, estrogen receptor [ER], progesterone receptor [PR], p53, and some CD markers) *(16)*. Increasing the sensitivity generally comes at a cost of decreasing specificity. The delicate balance that must be maintained requires a test battery approach to establish the optimal protocol for pH, time, and temperature that gives reliable, specific immunoreactivity *(8)*.

As always, standardization is a critical issue, especially for quantitative or semiquantitative immunohistochemical studies (Ki-67, ER, PR, p53). One area that has been particularly difficult to standardize is fixation. Which fixative is used and for how long has been virtually impossible to control across institutions. HIER has been used to reverse some of the effects of prolonged fixation. Another reagent that is of utmost importance is the antibody itself. Which epitope or part of an epitope does it recognize? Is it monoclonal or polyclonal? A number of antibodies that claim to recognize the same marker (for instance, CD3) are available with widely variable efficacy. Still another example is the use of antigen retrieval. As noted earlier, variations in time of retrieval, temperature, and pH of the buffer play a major role in the reproducibility and utility of these studies. In some instances, the use of AR requires re-evaluation of the clinical interpretation of IHC (e.g., wild-type p53 might be detected after using AR) *(17)*. The more cynical in the pathology community will often quip that any antigen they want to see as positive can be made so with the right preparation.

3.2.2. Detection Methods A wide variety of systems has been developed for the visualization of antigen–antibody reactions in tissue sections, and more are currently being investigated. Initially, direct staining methods using only a labeled primary antibody offered too little sensitivity. Secondary labeled antibodies directed against the primary antibody were introduced to increase the amount of signal generated by allowing more than one labeled antibody to bind for a given amount of antigen. Using an additional labeled antibody to react with the secondary antibody allowed further amplification.

The two most widely used antigen detection methods are (1) unlabeled, which include enzyme–antienzyme methods and (2) labeled, including the avidin–biotin complex (ABC), which eliminated the need to conjugate all primary antibodies *(6)*. Because of their low sensitivity and complex methodology, enzyme–antienzyme methods such as peroxidase–antiperoxidase (PAP) and alkaline phosphatase–antialkaline phosphatase (APAAP) techniques have largely been abandoned. The ABC and labeled streptavidin–biotin (LSAB) techniques have been used extensively. Newer detection systems are aimed at increasing sensitivity while limiting nonspecific background staining. Other practical goals are simplicity, reproducibility, and speed, all with less expense.

The ABC and LSAB techniques have had a significant role in the advancement of IHC and are still in use today *(18)*. ABC is based on the high-affinity binding between biotin and avidin. In this method, the secondary antibodies are conjugated with biotin and function as a link between the primary antibody, which is bound to the tissue, and the avidin–biotin complex. Streptavidin can be used instead of avidin with the advantage that there is less nonspecific tissue binding. A drawback of the ABC and LSAB methods is the fact that high levels of endogenous biotin in tissues, especially liver, kidney, thyroid, and brain, will produce background staining. Furthermore, HIER techniques appear to increase the reactivity of endogenous biotin to levels that can interfere with interpretation.

Newer, polymer-based amplifications are replacing the old three-step detection systems *(19–22)*. Depending on the method, either the primary or the secondary antibody is attached to an inert polymer (e.g., dextran) decorated with additional molecules of antibody and an enzyme such as horseradish peroxidase. Because the polymer can be bound to many antibody and enzyme molecules, the signal of an antigen with low-level expression can be greatly amplified. DakoCytomation (Carpenteria, CA) has pioneered this method by introducing the Enhanced Polymer One-Step Staining (EPOS) system. This resembles the early direct methods in that it requires only the primary antibody, with the difference of markedly increased signal in a single step. The methodology is easy, fast [might even be used for frozen section diagnosis *(23)*], and sensitive, but is limited by the number of labeled primary antibodies available. To address this problem, an indirect method that uses a standardized secondary antibody attached to dextran was developed. This EnVision system, from the same company, has been gaining popularity for its speed because of shorter incubation times, reduced complexity with only two steps, high sensitivity, and its ability to be used with a wide variety of primary antibodies *(20,21)*. Additional benefits include a potential reduction in primary antibody costs, because higher dilutions can be used while maintaining sensitivity, and in background staining, because endogenous biotin is avoided. The latter also allows for work with stronger HIER without concern for activation of endogenous biotin. A second generation of polymer-based detection reagents, PowerVision *(22)*, has been developed by ImmunoVision Technologies (Daly, CA). It uses small, multifunctional polymeric linkers that have better tissue penetration than dextran and, therefore, a higher sensitivity in detection.

Finally, signal amplification techniques such as tyramide amplification *(24,25)* can greatly enhance the sensitivity of classical methods such as ABC or LSAB. This technique is based on the generation of highly reactive biotinyl–tyramide intermediates in the vicinity of the peroxidase enzyme. Although this method can greatly increase the sensitivity of detection, nonspecific staining is also increased. Introduction of these methods requires careful evaluation of the interpretation criteria and of their clinical significance. For now, they remain primarily research tools.

3.2.3. Automation The fact that there are so many variables in IHC and that controlling these variables is essential to standardizing results has led to the development of automated

staining devices *(26)*. These devices reduce day-to-day staining variability and can potentially improve intralaboratory and interlaboratory reproducibility of results. There are many variations on the theme of automation. Most are based either on the use of capillary action or pipets to deliver reagents. They are devised as either closed systems that work only with proprietary reagent kits or more flexible open systems that can use antibodies and other reagents from different manufacturers. In general, the requirements for an automated staining device include analytical flexibility, "walk-away" operation, a user-friendly interface, and safety *(27)*. Taking automation to the next level has led to the production of automated instruments that will deparaffinize, perform antigen retrieval, and perform IHC staining. Automation can lead to more consistent quality by improving standardization, especially with regard to certain quantitative IHC stains that are potential prognostic indicators and therapeutic targets.

4. INTERPRETATION

Many areas of pathology are by their nature subjective, and IHC is no exception. What is considered positive will depend, to varying degrees, on the intensity of staining, the distribution of stain within the cell, and/or the percentage of cells showing staining. Ideally, the interpretation will be based on criteria set forth in the literature, but this is not always the case, especially because the literature frequently gives divergent data. The variation in published reports can be the result of differences in methodology, the specific antibody used, and, of course, the pathologist interpreting the slides *(28)*. A recent study of estrogen receptor staining in tissue microarrays of breast carcinomas documented significant interlaboratory variability in reporting of results. The variability was least in cases that were strongly positive, but increased dramatically in intermediate and negative cases. The study excluded fixation procedures and interobserver variability as major contributors. Variations in AR methods were cited as a main potential cause for disagreement *(29)*. Even with these attempts to standardize procedures, relatively little concerted effort has been made to standardize interpretation of results.

When evaluating IHC studies, simply knowing the cellular and subcellular localization of the protein being assayed can help eliminate some erroneous interpretations *(8)*. For example, cyclin D1 is a nuclear protein and one would not expect to find it in the cytoplasm (Fig. 1A). Similarly, the presence of HER-2/*neu* (HER-2) staining in the cytoplasm or nucleus should be regarded as nonspecific because its expression is restricted to the cell membrane (Fig. 1B). Comparison with appropriate internal positive and negative controls is essential and helps confirm that the antibody is marking the right cellular location. In cases where ambiguous results are obtained, validation can be attempted with antibodies against different epitopes of the same molecule or by detection of a related marker.

In addition to the lack of agreement on what constitutes "positive" at a cellular level, there is also the issue of what constitutes positive on a tissue level. In the breast, HER-2 protein overexpression and ER are prime examples. For HER-2 staining, the definition of "positive" rests on strong, uniform staining of the cell membrane, which has been proven by molecular studies to correlate well with amplification of the HER-2 gene. ER studies, on the other hand, have traditionally defined positive based on the percentage of cells that show nuclear reactivity. A cutoff of 10% cells with positive nuclear staining was generally accepted for reporting a positive result, although some argued for higher or lower percentages or for intensity of staining to be used. In 2000, the National Institutes of Health (NIH) issued a statement that any degree of positivity should be considered a positive result and a reason to offer antiestrogen therapy to breast cancer patients.

Finally, a number of artifacts can result in false-positive and false-negative results. In some instances, variable staining over a single slide begs the question of which area to believe. In such cases, either repeating the assay or finding an area in which the internal controls are appropriate might help resolve the issue. The absence of any staining of the tissue, in the presence of internal controls that should be positive, calls into question the immunoreactivity of the tissue or raises the possibility that some step in the process went awry. Testing with a ubiquitous marker such as vimentin can be of value in determining if the tissue has been rendered nonreactive because of processing, and a repeat assay will (hopefully!) solve the issue if the reagents were not appropriately applied.

4.1. CONTROLS Appropriate tissue controls known to react with the antibody being evaluated must be prepared and evaluated at the same time as the test sample. The presence of appropriately reactive positive and nonreactive negative internal controls is probably the best way to determine that an antibody was properly prepared and applied. However, some antigens are not expressed in normal tissues, and internal controls are absent. In these cases, it is necessary to include an external control tissue that is known to express the antigen being evaluated. Standardization of fixation is essential because differences in fixation between the control and test tissue might make a difference in the outcome of the staining procedure. For this reason, use of tissue processed "in house" as the control is most appropriate *(8)*. Because control and test slides are run at the same time, the issues relating to temperature, time, and pH of AR will be virtually identical. One issue that arises is whether the control should be placed on the same or a different slide. Ideally, having them on the same slide as the test material would afford more standardization, but this technique is more time-consuming and technically demanding. The choice of control material also has an impact on the interpretation of results. The control tissue should have the antigen of interest present at a detectable but not overpowering level. The use of low-level positive controls allows one to detect the antigen as well as detect slight alterations in primary antibody sensitivity. Surgical specimens should be used to prepare control slides to be used with routinely processed test specimens, whereas cytospin slides or cell block materials are more appropriate as controls for cytology specimens. In addition to positive controls, negative controls are necessary to evaluate nonspecific background staining or the presence of internal pigments that might interfere with interpretation. This is commonly accomplished by following the regular protocol, but with substitution of antibody diluent for the antibody of interest. Finally, in cases where there is little or no staining with the

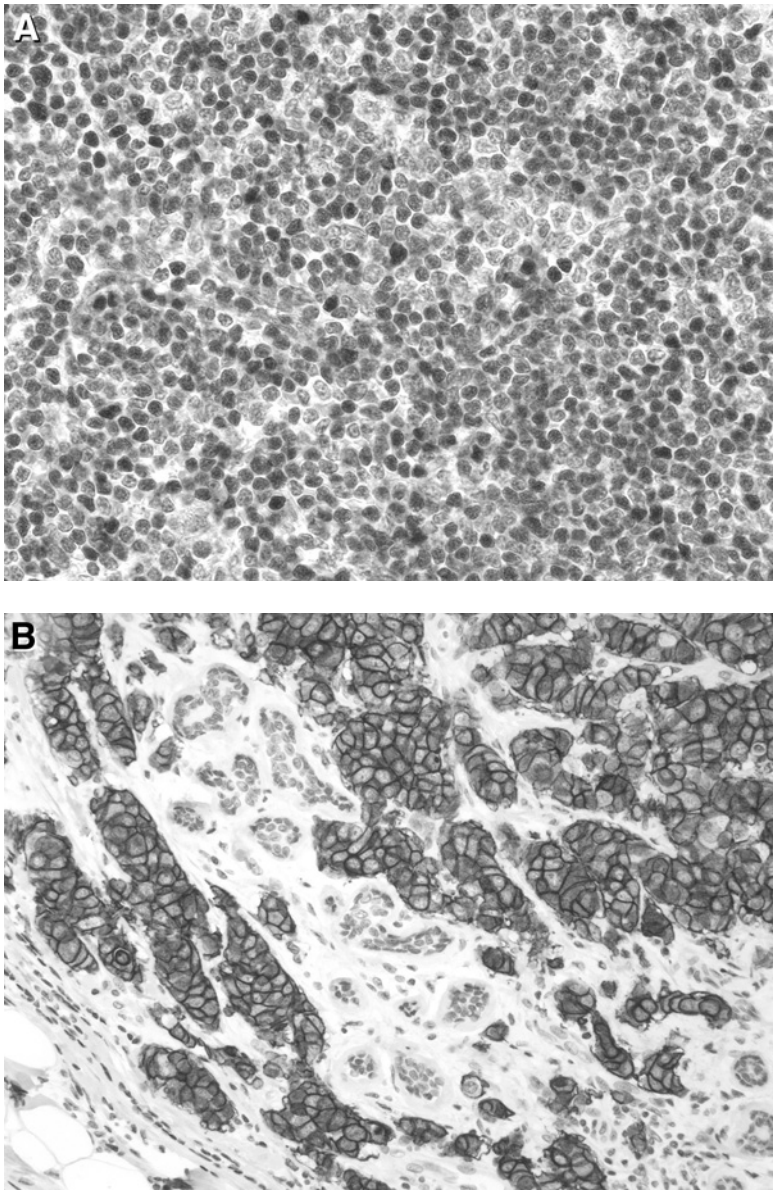


Fig. 1. (A) Nuclear staining for cyclin D1, characteristic for mantle cell lymphoma; **(B)** membranous staining for Her-2/*neu* (3+ positivity) in an invasive mammary carcinoma.

antibodies of interest, an antibody to a ubiquitous antigen such as vimentin might be used as a "reporter molecule" to ensure that the tissue was fixed and processed well enough to preserve immunoreactivity. For the present, controls are inherently imperfect because the amount of time between tissue removal and processing, fixation, and intensity can all vary from case to case and even within a given case. Suggestions to use specially cultured cells with more rigid control over these variables might be an improvement, but it will not address all concerns *(28)*.

4.2. APPLICATIONS

4.2.1. Immunohistochemistry of Infectious Disease

Infectious organisms sometimes induce cytopathic or histologic changes that are easily recognizable on hematoxylin and eosin (H&E)-stained sections or that can be highlighted by various histochemical stains such as Ziehl–Neelsen (for acid-fast bacilli [AFB]) or Grocott's methenamine silver (GMS) stain (for fungal organisms). In the case of GMS, the interpretation of positive versus negative is usually not difficult, as the organisms of interest are fairly large and visible at scanning magnification. On the other hand, searching for mycobacteria on an AFB stain can be an arduous task given the small size and scarcity of the organisms in the typical case. In addition to the ability to highlight the infectious agent, the specific identity of the organism can be elucidated by IHC *(30)*. There are antibodies available that recognize numerous infectious agents, of which antibodies against viral antigens such as CMV and HSV and bacterial antigens associated with *Helicobacter pylori* are among the most commonly used. With the appropriate specimen preparation and control material, these tools can increase the sensitivity and specificity of tissue examination for infectious agents. Although the degree of antibody specificity varies depending on the organism, the turnaround time on IHC is often far faster than

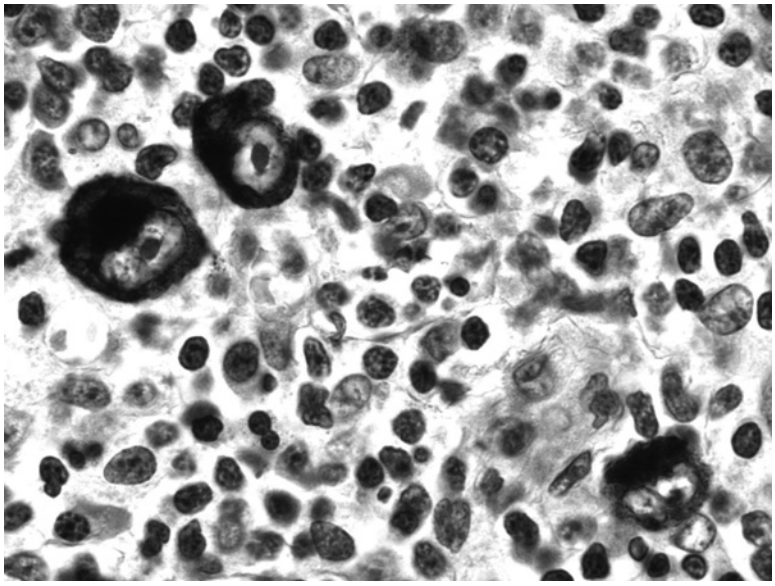


Fig. 2. Membranous and cytoplasmic staining for EBV-LMP1 in neoplastic cells from a case of Hodgkin's lymphoma.

cultures or molecular tests. Markers for certain fungal organisms, such as *Cryptococcus neoformans,* are quite specific, showing virtually no crossreactivity with other fungi. On the other hand, the current iterations of antibodies that recognize *Mycobacterium tuberculosis* (MTB) show crossreactivity with other mycobacterial species such as *Mycobacterium–avium intracellulare* complex. Likewise, IHC for *H. pylori* shows cross reactivity with the closely related *H. heilmannii*, although the morphology of these two organisms is sufficiently distinct to allow differentiation. Molecular studies available for mycobacteria and for *H. pylori* do not have these issues of crossreactivity, but are more time-consuming and expensive than IHC. A culture of these micro-organisms can take weeks to months, whereas IHC or molecular studies for MTB can provide results in a matter of days.

4.2.2. Detection of Oncogenic Viruses Epstein–Barr virus (EBV) is one member of the herpesvirus family that has oncogenic potential. It has been linked to the development of a variety of lymphoid and epithelial malignancies with varying frequencies *(31,32)*. It is virtually always associated with nasopharyngeal carcinoma, commonly seen in posttransplant lymphoproliferative disorder (PTLD) and less frequently in Burkitt lymphoma and Hodgkin lymphoma. More recently, an association has been suggested with gastric and breast carcinomas *(32,33)* although the latter is still a matter of debate *(33)*. Although the exact mechanisms of malignant transformation are yet to be elucidated, the detection of viral nucleic acids or proteins can serve as a useful marker for diagnosis in some cases. EBV can persist as a latent infection in non-neoplastic B-cells and is also present in the latent form in the EBV-associated tumors. The latently infected cells express a limited number of viral genes, coding for six EBV nuclear antigens (EBNAs), three latent membrane proteins (LMPs), and early RNAs (EBERs). Interestingly, different tumors show consistent patterns of EBV gene expression and can be categorized as latency types I, II, and III based on the pattern of expression of these genes *(34)*.

There are two main methods of detecting EBV in tissue: EBER *in situ* hybridization and immunohistochemical detection of EBV latent proteins *(35)*. Currently, detection of EBER by *in situ* hybridization is regarded as the "gold standard" for detection and localization of EBV in tissues. EBERs are consistently expressed at high levels in all EBV-associated tumors, as well as in the lymphoid tissues of patients with acute infection (infectious mononucleosis). Development of antibodies against viral proteins has permitted IHC detection of EBV-infected cells. Antibodies against LMP1 are currently used to detect some EBV-associated tumors (Fig. 2), but LMP1 expression is seen only in latency type II (Hodgkin lymphoma and nasopharyngeal carcinoma) and type III (PTLD) tumors, not in latency type I (Burkitt lymphoma) tumors. The development of antibodies against other EBV proteins such as EBNA1, EBNA2, and LMP2 might help further characterize these tumors. For instance, all EBNAs and LMPs are expressed in latency type III, whereas latency type II tumors express EBNA1 and the three LMPs, and Burkitt lymphoma expresses only EBNA1.

Other viruses implicated in the development of human neoplasms include human papilloma viruses (HPVs), which are associated with squamous cell carcinoma of the cervix, and human herpesvirus 8 (HHV-8 or KSV), which is associated with Kaposi sarcoma (KS), body cavity lymphomas, as well as multicentric Castleman disease. At least one study has demonstrated that antibodies against HHV-8 latent nuclear antigen detect the virus in cases of KS with good correlation with polymerase chain reaction (PCR) methods. *(36)*. On the other hand, detection of HPV in Pap smears by IHC has focused on looking at p16, a cyclin-dependent kinase inhibitor whose expression is upregulated by HPV in squamous and glandular cervical neoplasms *(37,38)*.

4.3. APPLICATIONS OF IMMUNOHISTOCHEMISTRY TO TUMOR DIAGNOSIS Before the era of modern laboratory tests such as molecular pathology and IHC, pathology diagnoses were made on the basis of H&E morphology, a few

histochemical stains, and clinical data. It is well recognized that morphology can be misleading, as there is considerable overlap in the appearance of various disease entities. In tumor pathology, the examples often given are renal cell carcinoma and melanoma as the great morphologic mimics. With a panel of immunohistochemical markers, these tumors can now usually be diagnosed with confidence. IHC has greatly enhanced the accuracy of tumor diagnosis *(39),* particularly in the diagnosis of poorly differentiated tumors and lymphomas, defining the origins of metastatic tumors of unknown primary and spindle cell neoplasms. The contributions of IHC to resolving these problems with some specific examples are illustrated in the following subsections.

4.3.1. Diagnosis of an Undifferentiated Malignant Neoplasm It is not uncommon to encounter undifferentiated malignant neoplasms whose H&E appearance gives few clues as to its origin. In the past, such tumors might have been signed out as "poorly differentiated malignant neoplasm," which is a frustrating diagnosis for the clinician, pathologist, and patient alike. Immunohistochemical analysis using a limited panel of antibodies such as a "cytokeratin cocktail" (mixture of various high- and low-molecular-weight cytokeratins), S-100 and/or human melanoma black (HMB)-45, and the leukocyte common antigen (LCA or CD45) can help establish the lineage of the tumor as epithelial, melanocytic, or hematopoietic. The ability to differentiate melanoma, carcinoma, and hematologic malignancies is essential to appropriate patient management. Additional studies might further subclassify these entities to help the clinician more precisely tailor therapy.

4.3.2. Establish the Site of Origin in Metastatic Carcinoma of Unknown Primary In other instances, metastatic disease is the first or only manifestation of a carcinoma. If a primary is not detected clinically, it is up to the pathologist to suggest a potential site of origin for the neoplastic process. Again, IHC can help detect the site of origin in cases without a clinically recognizable primary *(40)*. There are a few markers that suggest a particular primary site such as prostate-specific antigen (PSA) in prostate, thyroglobulin in thyroid neoplasms, gross cystic disease fluid protein (GCDFP-15) and ER/PR in breast cancer, HepPar-1 in hepatocellular carcinoma, and thyroid transcription factor (TTF)-1 for lung or thyroid primaries. Even though these markers give clues to the origin of the tumor, none is diagnostic of a particular primary site. In some cases, when an antibody is introduced as a site- or type-specific marker, the passage of time and experience with more tumor types shows some immunohistochemical overlap with unrelated tumors. For this reason, interpretation of these markers should be done with foreknowledge of possible pitfalls. Because most tumors do not have a pathognomonic immunohistochemical marker, one way to help narrow the field of possibilities is to use combinations of antibodies to create an antigen expression profile that will increase or decrease the likelihood of certain primary sites *(40,41)*. Such a panel might include various low- and high-molecular-weight cytokeratins, CK7 and CK20, carcinoembryonic antigen (CEA), and possibly TTF-1 (Table 1).

Table 1
IHC Panel in Carcinoma of Unknown Origin

Tumor	*CK7*	*CK20*	*CEA*	*TTF-1*
Small-cell cacinoma	–/+	–	–/+	+
Lung adenocarcinoma	+	–	+	+/–
Hepatic	–	–	–	–
Breast	+	–	–	–
Endometrium	+	–	–	–
Stomach	+/–	+/-	+	–
Pancreas	+/–	–/+	+	–
Colon	–	+	+	–
Prostate	–/+	–	–	–
Renal	–	–	–	–
Thyroid	-/+	+/–	–	+

4.3.3. Lymphoma Diagnosis In some fields, notably hematopathology, immunophenotype is an essential component of classification of the various neoplasms *(42,43)*. Immunophenotypic studies are an integral part of lymphoma diagnosis, as clearly stated in the recent WHO classification of hematolymphoid tumors. Immunophenotype can be determined by flow cytometric analysis if fresh tissue is available or by IHC on routinely processed tissue. Because of its quantitative properties, the first technique can be invaluable in finding subtle clonal B-cell populations and detecting aberrant T-cell antigen expression. In addition, some markers that are detected on the fresh cells used for flow cytometry (e.g., the B-cell marker CD19) are not yet available for use on formalin-fixed tissue. However, more and more, lymphoid markers are becoming available for use in paraffin-embedded tissue. The unique ability of IHC to combine morphology and immunophenotype is one reason that it is currently the method of choice compared to other tests such as flow cytometry and molecular studies in the evaluation of lymphoid neoplasms.

Some B-cell lymphomas can be accurately diagnosed based on morphologic grounds, but the presence of morphologic overlap between different entities usually requires IHC studies using a variety of markers. CD20 is a reliable B-cell marker with membranous staining that is positive in the large majority of small and large B-cell lymphomas. Pax-5, a more recently introduced B-cell marker, is localized in the nucleus of normal and neoplastic B-cells. It is positive in B-cell lymphomas (including precursor B-cell lymphoblastic lymphoma, which is usually negative for CD20) as well as in Hodgkin lymphoma *(44)*. Although this marker is gaining in popularity, the extensive experience with staining and interpretation of CD20 as well as the lower cost make it the antigen of choice for characterizing lymphoid cells as B-cells. In some special circumstances, such as patients receiving anti-CD20 therapy (rituximab), Pax-5, or CD79a can help demonstrate B-cell lineage. Pax-5 is also of value in demonstrating B-lineage of precursor B-cell neoplasms and in distinguishing Hodgkin lymphoma from anaplastic large-cell lymphoma. The antiapoptosis protein bcl-2 is a marker that distinguishes follicular hyperplasia (bcl-2-negative) from follicular lymphoma that typically overexpresses bcl-2 *(45)*. It is a common misconception that the presence of bcl-2-positivity supports that a lymphoma is of follicle center origin. This is not true, because a variety of neoplastic and non-neoplastic lymphoid cells express this antigen.

Table 2
IHC Panel in Small B-Cell Lymphomas

Lymphoma	*CD5*	*CD23*	*CD43*	*Cyclin D1*	*BCL6*
FCC	–	–	–	–	+
MCL	+	–	+	+	–
MZL	–	–	+/–	–	–
CLL	+	+	+	–	–

Abbreviations: FCC, follicular lymphoma; MCL, mantle cell lymphoma; MZL, marginal zone lymphoma; CLL, chronic lymphocytic leukemia.

The use of panels of markers (CD5, CD23, CD43, cyclin D1, CD10, bcl-6) are used to subclassify the small B-cell lymphomas *(46)*. Some, such as small lymphocytic lymphoma, mantle cell lymphoma, and follicular, have a characteristic immunophenotype (Table 2). For instance, among these markers, the presence of cyclin D1 is considered specific for mantle cell lymphoma *(47)*. Markers such as Ki-67, bcl-6, and CD10 are being investigated as possible prognostic markers in large-B-cell lymphomas *(48,49)*.

Among the most valuable and sometimes the most frustrating IHC markers in the study of B-cell processes are the immunoglobulin light-chain studies. The ability to demonstrate immunoglobulin light-chain restriction is one of the most specific methods to identify neoplastic B-cells. Light-chain restriction can be reliably identified by flow cytometry, if fresh tissue is available. Light-chain restriction can also be easily demonstrated by IHC in plasma cell neoplasms and lymphomas with plasmacytic differentiation. However, inconsistencies and technical problems have beset the identification of light-chain restriction for most other B-cell neoplasms. More recently, studies have shown more consistent results with a modified technique using HIER *(50,51)*.

A series of T-cell markers (CD2, CD3, CD4, CD5, and CD8) is also available for use on paraffin-embedded tissue. Because there is not a good analog to the immunoglobulin light-chain studies in B-cell processes, one must rely on loss of expression of various T-cell markers, altered intensity of expression in the neoplastic cells compared to normal, or alteration of the CD4 and CD8 staining (either marked predominance of one or the other, or T-cells that express both or neither marker). These changes can be suggestive of a T-cell lymphoma, but the final diagnosis of T-cell lymphoma in many cases is dependent on molecular studies to confirm a clonal rearrangement of the T-cell receptor gene *(52)*. As with the B-cell neoplasms, a few T/NK (natural killer)-cell neoplasms, such as anaplastic large-cell lymphoma, have distinct IHC and molecular features.

Hodgkin lymphoma (HL) comprises two distinct groups of diseases: classical HL and nodular lymphocyte predominant HL (NLPHL). Both are characterized by relatively small numbers of neoplastic cells in a background of mixed inflammation in the case of classical HL or lymphocytes in the case of NLPHL. Cells morphologically identical to the large binuclear and mononuclear Reed–Sternberg (RS) cells that typify classical HL can be seen in NLPHL, as well as many forms of non-Hodgkin lymphomas (NHL). Likewise, cells virtually identical to the L&H cells of NLPHL can be identified in other conditions. Immunohistochemistry is essential in distinguishing the two types of Hodgkin lymphoma: the RS cells in classical HL are CD30 and CD15 positive, whereas the L&H cells in NLPHD are negative for CD15 and CD30 and express CD20. In fact, L&H cells not only show a B-cell phenotype but also show clonal immunoglobulin gene rearrangements *(53)*. With improvements in IHC techniques, even the "classic." immunophenotype of RS is being reevaluated because many cases of HL show at least some CD20-positivity in the neoplastic cells. Additional markers such as BOB.1 and Oct2 *(54)* or Pax-5 *(44)*, although not 100% specific, can be added to a panel of markers that help distinguish NLPHL from classical HL, or classical HL from its imitators. The immunophenotype of RS cells in a given case greatly depends on technical factors such as fixation and AR *(55)*. Although clonality can be demonstrated in the RS cells with microdissection techniques, the relatively low number of neoplastic cells in cases of HL, means that any clonal bands by PCR would likely to be obscured by reactive cells in the background.

4.3.4. Sarcoma Diagnosis Soft tissue tumors can show a wide variety of histologic appearances, and morphologic overlap is often the rule, rather than the exception. Although there are few specific markers available, panels of antibodies can frequently be used to identify the cell of origin. Some relatively specific markers, including myogenin (skeletal muscle tumors), CD31 (vascular tumors), and CD117 (gastrointestinal stromal tumors), exist, but others, like S-100, react not only with neural-derived tumors but also other tumor types such as melanoma. The coexpression of cytokeratin by a limited number of sarcomas such as synovial sarcoma, chordoma, and epithelioid sarcoma can be a useful feature in subtyping soft tissue tumors, but this can lead to confusion with carcinoma in some cases.

Significant progress has been made in the understanding of the molecular pathogenesis as well as immunohistochemical characteristics of the "small blue cell tumors" of childhood. These include sarcomas such as rhabdomyosarcoma, Ewing sarcoma/primitive neuroectodermal tumor (ES/PNET), desmoplastic round cell tumor (DRCT), as well as neuroblastoma and acute lymphoblastic lymphomas (ALL). Morphologic overlap is very common among these tumors, whereas prognosis and therapy are becoming more individualized. Molecular pathology has elucidated characteristic translocations in some of these sarcomas [alveolar rhabdomyosarcoma *(56)*, ES/PNET *(57)*, DRCT *(58)*] and molecular techniques such as reverse transcription (RT)–PCR are able to identify these abnormalities even in formalin-fixed, paraffin-embedded tissue *(59)*. However, these techniques are available only in specialized centers, and the majority of pathologists rely on IHC panels for the diagnosis (Table 3).

In the case of markers of cell lineage, desmin, myoD1, and myogenin indicate muscle differentiation *(60)*. Desmin is a highly sensitive marker for rhabdomyosarcoma, but it lacks specificity. MyoD1 and myogenin are muscle-restricted nuclear transcription factors, highly specific for skeletal muscle differentiation and, therefore, for rhabdomyosarcomas *(61,62)*. Among them, myogenin is the preferred marker (Fig. 3) because myoD1 usually shows more nonspecific background staining.

The ES/PNETs are positive for CD99 and for the more recently introduced marker FLI-1. Although CD99, the product

Table 3
IHC Panel in Small Blue Cell Tumors

	Desmin	*Myogenin*	*CD99*	*CD43*	*NF*	*FLI-1*	*WT-1*
RMS	+	+	–/+	–	–	–	–
EW	–	–	+	–	–	+/–	–
DRCT	+	–	–	–	–	–	+
NB	–	–	–	–	+	–	–
LL	–	–	+/–	+	–	–/+	–

Abbreviations: MS, rhabdomyosarcoma; EW, Ewing sarcoma; DRCT, desmoplastic small round cell tumor; NB, neuroblastoma; LL, lymphoblastic lymphoma.

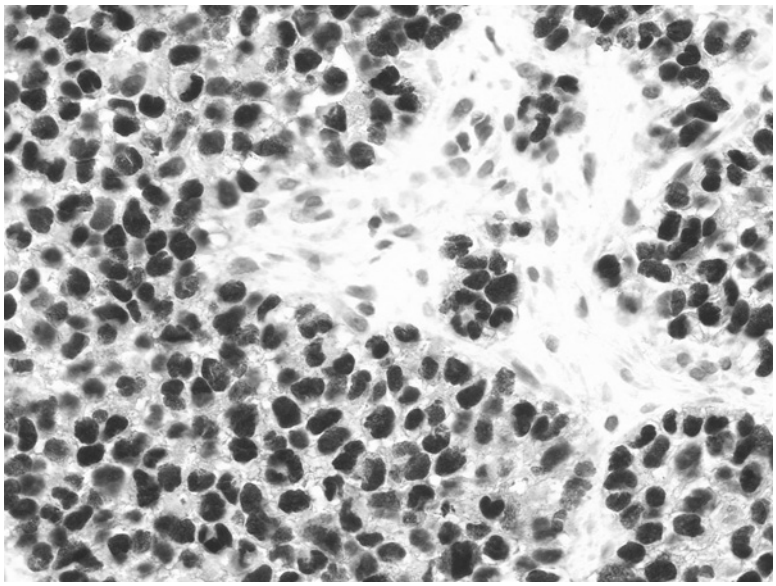


Fig. 3. Diffuse nuclear staining for myogenin in alveolar rhabdomyosarcoma.

of the MIC2 gene, is very sensitive for ES/PNET *(63)*, it is not as specific as it was initially thought because it has been documented in other small-cell sarcomas such as rhabdomyosarcoma, synovial sarcoma, and ALL. Recently, overexpression of the nuclear protein FLI-1 has been demonstrated in about 70% of ES/PNET *(64)*. The majority of ES/PNET cases show the translocation t(11;22)(q24;q12) that leads to production of the EWS-FLI-1 fusion protein, which is a transcription activator with transforming potential. Whereas the EWS/FLI-1 chimeric gene is specific for ES/PNET, the IHC antibody that recognizes FLI-1 is not. Positive staining for FLI-1 has not been seen in other small-cell sarcomas, but has been documented in ALL and some vascular tumors.

Another small round blue cell tumor with a distinctive phenotype is DRCT. It shows coexpression of cytokeratin, vimentin, desmin, and neuron-specific enolase *(65)*. This tumor has a characteristic translocation, t(11;22)(q13;q12), which gives rise to the production of the chimeric EWS/WT-1 protein. Antibodies against WT-1 are used to detect this protein by IHC *(66)*. This protein is not entirely unique to DRCT as it is also expressed in Wilms tumor. However, among the other small round blue cell tumors, its presence is specific for DRCT.

4.3.5. Tumor Prognostic Factors Although IHC is predominantly used to determine tumor lineage or highlight tumor cells, it is becoming increasingly evident that some IHC markers have prognostic value and even direct therapeutic implications *(67)*. Among these are ER/PR and HER-2, which are routinely evaluated by IHC to help direct therapy in breast carcinoma. The expression of CD117 (c-Kit) helps separate gastrointestinal stromal tumors (GISTs), which respond well to imatinib, a tyrosine kinase inhibitor, from other mesenchymal tumors of the gastrointestinal (GI) tract.

p53 expression is another commonly used prognostic marker *(68,69)*. The *p53* gene, located on chromosome 17p, encodes a nuclear DNA-binding protein that regulates cell growth. Mutant forms of the p53 protein are resistant to degradation and accumulate in the nucleus, where it is detectable by IHC. In the normal state, p53 is rapidly degraded and is usually not detectable. Overexpression or accumulation of this protein is associated with a poor prognosis in a wide variety of tumors. Studies have also shown that tumors harboring mutations in p53 might be more resistant to chemotherapy and radiation therapy.

4.3.6. Tissue Microarrays A significant technical advancement is the recent introduction of the tissue microarray (TMA) method. This technique involves harvesting small 0.6-mm cores of tissue from hundreds of archival paraffin blocks and constructing a single TMA block, which can be used not only for IHC but also for *in situ* hybridization techniques. With this technique, one can examine many examples of a given tumor or group of tumors at one time using a minimum of resources. It is a highly efficient method to that allows for

fast, parallel immunohistochemical profiling of large numbers of cases and is an excellent way to evaluate new protein markers. It can serve as a powerful quality assurance tool in assessing the specificity and sensitivity of different antibodies *(70)*. In addition, its use in the research laboratory might contribute to the expeditious translation of new molecular findings into clinical applications.

4.4. MOLECULAR/IMMUNOHISTOCHEMICAL CORRELATION The following are common examples in which genetic events translate into phenotypic changes that can be detected by immunohistochemistry *(71–73)*. In some cases, these changes are best detected by molecular means such as fluorescence *in situ* hybridization (FISH) or PCR, whereas in others, IHC is the method of choice. In many instances, the use of IHC is simply an easier and more accessible laboratory test that gives virtually the same information. In other areas, IHC serves well as a screening test, whereas the PCR or *in situ* hybridization studies play more of a confirmatory role.

4.4.1. Genetic Alteration Resulting in Overexpression of a Protein

4.4.1.1. Her-2/Neu in Breast Carcinoma The field of IHC has been fruitful in the search for independent prognostic indicators in breast carcinoma. The detection of ER and/or PR in the tumor cells predicts response to hormone-related therapy and a better outcome. In contrast, overexpression of HER-2, an oncogene in the epidermal growth factor receptor family, has been associated with poorer outcome with respect to disease-free status and overall survival *(74)*. On the other hand, some studies have shown that such patients benefit from doxorubicin therapy and the humanized monoclonal antibody to the HER-2 protein, trastuzumab. Although there is agreement that this is an important parameter to be assessed in patients with breast carcinoma, there is considerable debate as to the most appropriate way to evaluate tumors for this overexpression.

Although immunohistochemistry is a quick, inexpensive method for evaluation of HER-2 overexpression, a lack of standardization of techniques, use of multiple antibodies, and variable scoring have plagued the assessment of HER-2 by IHC *(75)*. With standardized protocols for the assay and the evaluation of results, reproducibility is much improved. A number of preanalytic factors can alter results, including difference in type and time of fixation and selection of appropriate control and test material. The scoring of HER-2 IHC ranges from 0 (no membrane staining in invasive tumor) to 3+ (strong, complete membrane staining) *(76)*. Practice with the standard scoring system should reduce variation. Most studies demonstrate a close correlation between strong positivity (3+) by IHC (Fig. 1B) and amplification of the HER-2 gene by FISH studies. There is less agreement when the IHC results are intermediate, although the large majority of 1+ and 2+ results by IHC are shown to be negative for amplification by FISH. FISH and, more recently, chromogenic *in situ* hybridization (CISH) appear to work well in archived formalin-fixed tissue. These assays are typically more costly and time-consuming than IHC, but the level of variability seen in IHC studies has prompted some to advocate FISH as the method of choice for screening patients for Her2/neu overexpression *(77)*. An advantage of FISH is that it allows one to determine in part the mechanism by which Her2/*neu* is overexpressed. It has been shown that the majority of cases of overexpression are caused by gene amplification, and this technique permits the observer to quantify that amplification by actually counting the number of copies of the Her2/*neu* gene present in the tumor cells. Processing with enzymatic digestion can lead to either no signal or to autofluorescence that might obscure signal, leading to reports in the literature of false-negative rates up to 10%. In addition, slide preparation leads to distortion of tissue morphology such that it might be difficult to distinguish between invasive and *in situ* components of tumor.

A recent symposium offered by the College of American Pathologists offered a "consensus testing algorithm" *(78)* for evaluation of Her2/neu status in breast carcinomas. They suggest that if a lab's concordance between FISH and IHC studies is high enough (>90% for scores of 3+ and 0), then IHC can be used as a screening test with reflex testing by FISH for cases that show 1+ or 2+ reactivity by IHC. If the concordance rates are lower than 90%, the laboratory should consider running all tests by FISH.

4.4.1.2. c-Kit Overexpression in Gastrointestinal Stromal Tumors The gastrointestinal stromal tumor (GIST) is the most common mesenchymal neoplasm of the gastrointestinal (GI) tract. This entity has been a source of confusion in terms of classification for many years because it has an ability to show differentiation along several different lines such as myogenic (smooth muscle), neural, and bidirectional differentiation. It was shown that GISTs originate from the interstitial cells of Cajal, also known as GI pacemaker cells, which have immunophenotypic and ultrastructural characteristics of both smooth muscle and neuronal differentiation in varying degrees and serve to regulate peristalsis. They are characterized by expression of c-Kit (CD117) in virtually all cases (Fig. 4), CD34 in over 70%, smooth muscle actin in less than 30%, and, less commonly, S-100 protein and desmin *(79)*.

c-Kit is a transmembrane tyrosine kinase receptor that is expressed at high levels in hematopoietic stem cells, mast cells, melanocytic cells, germ cells, and the interstitial cells of Cajal (ICC). In 1998, Hirota and colleges reported that some GISTs contain an exon 11 mutation in the c-Kit proto-oncogene *(80)*. This mutation in the c-Kit gene leads to overexpression of the tyrosine kinase moiety for the c-kit protein (CD117), which is recognized as a reliable phenotypic marker for this neoplasm. Mutations that result in c-Kit overexpression are thought to play a major role in the pathogenesis of GIST. The recent introduction of a receptor tyrosin kinase inhibitor (STI-571, imatiniber or Gleevec) that inhibits activated c-Kit protein, for the first time provided an effective treatment for recurrent or metastatic GIST *(81)*. Expression of CD117 is a reproducible diagnostic criterion for GIST and was a requirement for many clinical trials involving STI-571, although recent data suggest that even in the absence of c-Kit staining by IHC, GISTs might respond to STI-571. It has been proposed that the term GIST be applied only to tumors expressing CD117, although rare exceptions might exist that include lesions that appear immunohistologically nonreactive secondary to poor fixation and preparation, are c-Kit negative because of sampling error, or have ceased to express c-Kit because of some form of clonal

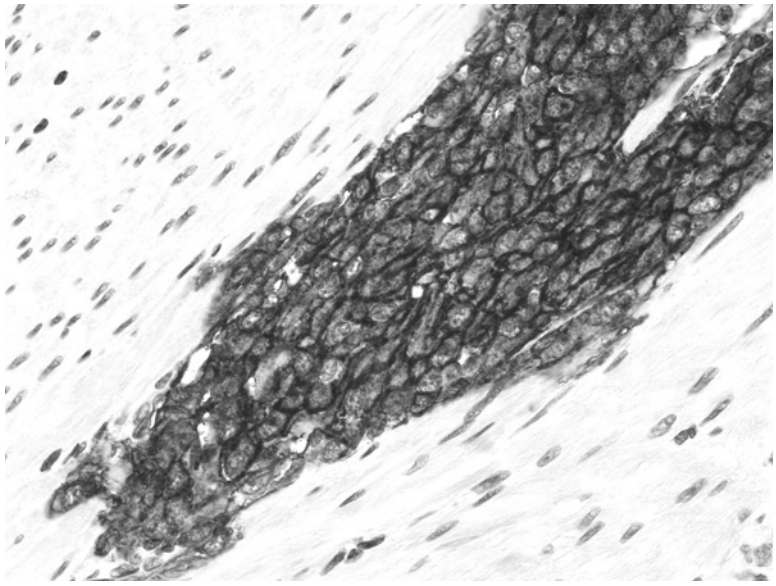


Fig. 4. Diffuse, intense staining for c-Kit (CD117) in GIST.

evolution (STI-571 therapy). Antibodies to c-Kit typically show diffuse cytoplasmic and/or membranous staining, with rare tumors showing perinuclear staining *(82)*. Mast cells serve well as a positive internal control to confirm immunoreactivity of the tissue and that the stain is working properly. In less than 2% of cases, an otherwise typical tumor lacks c-Kit overexpression and should probably be labeled as spindle cell (or epithelioid) stromal neoplasm most consistent with GIST *(80)*.

It should be noted that other tumors that involve the GI tract might also express c-kit. Melanoma, seminoma, granulocytic sarcoma, as well as other malignant spindle cell tumors have been reported to be positive for CD117. For this reason, the expression of CD117 must always be interpreted in the context of the H&E morphology and clinical findings.

4.4.2. Genetic Mutations Resulting in Loss of Protein Expression

4.4.2.1. E-cadherin in Lobular Breast Carcinoma E-Cadherin is a cell surface glycoprotein involved in cell adhesion. The protein is encoded by the CAD1 gene on chromosome 16q and loss of its expression is associated with increased invasiveness and higher-tumor grade *(83)*. Dysregulation of E-cadherin is seen in many carcinomas and is caused by a heterogenous mix of genetic mutations that can be detected by molecular methods *(84)*. Mutations of the E-cadherin gene are particularly prevalent in poorly cohesive neoplasms like lobular carcinoma of breast and gastric carcinoma *(85)*. Of interest, mutations of this gene are rarely seen in ductal or medullary carcinoma of the breast. Because of this dichotomy, IHC can be used to separate lobular and ductal breast carcinomas that have ambiguous H&E morphology. Lack of staining with antibodies to E-cadherin in an *in situ* or invasive breast carcinoma indicates lobular rather than ductal carcinoma.

4.4.3. Mismatch Repair Gene Expression in Colonic Carcinoma

There are two recognized genetic pathways for the development of colonic carcinoma. The less common of the two is seen in about 10–15% of sporadic cases of colon cancer and all cases of hereditary nonpolyposis colon cancer syndrome (HNPCC) or Lynch syndrome *(86)*. Over 90% of these cases show mutations in the hMLH1 and hMSH2 genes whose products are involved in DNA mismatch repair, a critical proofreading function in DNA replication. A germline defect in one of these genes carries with it a lifetime risk for colon cancer of about 80%. The result of such alterations is microsatellite instability (MSI), which is the molecular hallmark of this group of tumors. The clinical relevance for identifying this group of patients is that they have a better survival rate but have a higher incidence of metachronous tumors. In addition, it allows for earlier screening of potentially affected relatives. Although molecular studies for alterations in these genes are available, tumors can be screened in a technically easier and less costly way with IHC for hMLH1 or hMSH2 proteins. Loss of expression of these gene products is an effective surrogate for the molecular presence of microsatellite instability *(87,88)*.

4.4.4. Translocations Result in Expression of a Chimeric Protein or Activation of *a* Normal Gene with Overexpression of a Normal Protein

4.4.4.1. ALK1 in Anaplastic Large Cell Lymphoma One entity that has been defined by IHC is the CD30-positive anaplastic large-cell lymphoma (ALCL) *(89)*. The classic form of ALCL has sheets of large "hallmark" cells with vesicular embryoid or reniform nuclei, one or multiple nucleoli, abundant amphophilic cytoplasm, and a perinuclear hof, Although many morphologic subtypes exist, they have in common the expression of CD30 in a membrane and Golgi distribution (Fig. 5A). A subset of these tumors were shown to have a characteristic t(2;5)(p23;q35) chromosomal translocation that fuses the ALK and NPM genes and leads to production of a novel NPM–ALK fusion protein containing the N-terminal portion of nucleophosmin and the cytoplasmic domain of ALK, a neural-associated receptor tyrosine kinase *(90)*. It has not been possible to predict which tumors will have abnormal ALK expression on the basis of morphology.

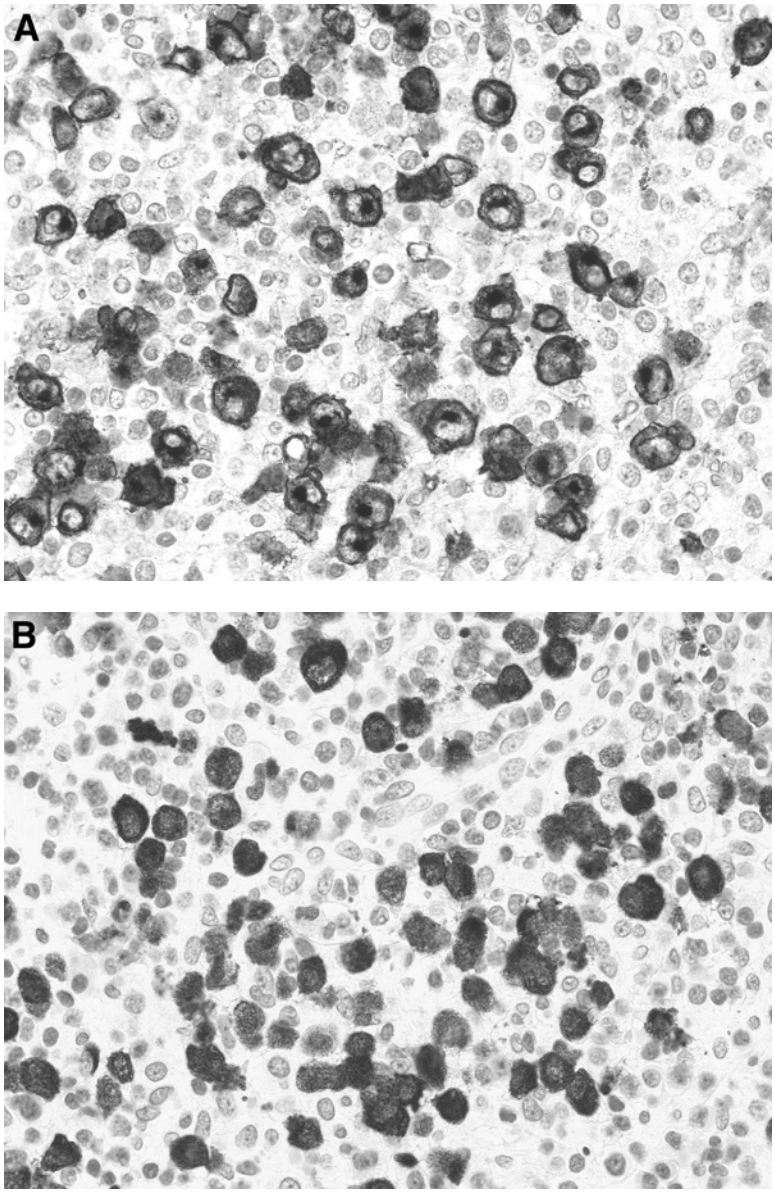


Fig. 5. (A) CD30 highlights neoplastic cells in ALCL; **(B)** ALK-positive neoplastic cells in ALCL, with cytoplasmic and nuclear staining.

The laboratory that initially described the p80 protein (NPM–ALK) first raised polyclonal antibodies to the tyrosine kinase domain *(91)*. Since that time, purified monoclonal antibodies have been produced that are used for routine IHC analysis. This marker is quite specific in that it marks no other type of lymphoma and is absent from normal tissues except for weak positivity in a small subset of cells in the brain. A few nonlymphoid entities have been shown to be positive for ALK1, including inflammatory myofibroblastic tumor and rare soft tissue tumors. In cases of ALCL, the IHC staining pattern gives an idea about the underlying translocation involved. The classic NPM–ALK fusion protein shows pattern of nuclear and cytoplasmic staining because although ALK is present mainly in the cytoplasm, wild-type NPM and the NPM–ALK fusion protein can form dimers localized to the nucleus (Fig. 5B). When translocation partners other than NPM are present, the staining is typically limited to cytoplasm or to the cell membrane. ALK expression by the neoplastic cells is a favorable prognostic factor, but the specific genetic rearrangement leading to that overexpression does not seem to correlate with outcome *(92)*.

The translocations in ALCL occur over large areas within the ALK and NPM, thereby limiting the use of standard PCR as a diagnostic tool. Furthermore, up to 20% of cases have translocation partners other than NPM, which would not be detected even with RT-PCR. FISH studies can detect either the specific t(2;5) or can evaluate for ALK rearrangements, allowing detection of variant translocations involving the ALK gene. Given the fact that IHC is less expensive, faster, and easier than the molecular methods and the fact that the pattern of staining can predict variant translocations, IHC testing is favored for evaluation of ALCL *(93)*.

4.4.4.2. Cyclin D1 in Mantle Cell Lymphoma Overexpression of cyclin D1, a nuclear protein whose expression permits the cell to transition from the G_1-phase to the

S-phase in the cell cycle, is a defining feature of mantle cell lymphoma. In this lymphoma, a characteristic chromosomal rearrangement, t(11;14)(q13;q32), juxtaposes the bcl-1 (CCND1 gene) locus at 11q13 with the IgH locus at 14q32. Although increased cyclin D1 expression is the hallmark of mantle cell lymphoma, it is also reported in multiple myeloma and hairy cell leukemia *(72)*. Detection of cyclin D1 by IHC has been hindered by technical difficulties *(94)*. In some instances, cyclin D1 staining can show nonspecific cytoplasmic staining (caused by biotin) that is clearly distinct from the mosaic pattern of nuclear staining that characterizes a positive reaction (Fig. 1A). In negative cases, one can assess the internal control seen as positive nuclear staining in scattered endothelial cells. With careful preparation using a high-temperature, high-pH AR method, it can be detected in over 95% of cases of mantle cell lymphoma *(95)*. By contrast, PCR methods on paraffin-embedded tissue detect the t(11;14) rearrangement in 50–70% of cases of mantle cell lymphoma. More recent reports using FISH give a detection rate close to 100% *(96,97)*. This FISH procedure can be applied to formalin-fixed, paraffin-embedded tissue as well as unfixed cells.

Similar mechanisms of overexpression of regulatory proteins produce other lymphomas. For instance, in follicular lymphoma, the antiapoptosis protein bcl-2 is commonly overexpressed because of translocation of the bcl-2 gene on chromosome 18 is translocated to the IgH locus on chromosome 14, leading to overexpression. This overexpression blocks normal apoptosis, leading to overgrowth of follicle center cells.

5. SUMMARY

Immunohistochemistry and molecular diagnostics have been developing in parallel for some time. Because IHC is more readily accessible and its interpretation is more akin to routine morphologic examination of tissues, it has become the favored method for fine-tuning diagnoses in surgical pathology. As demonstrated here, in many instances IHC serves as a surrogate for molecular (i.e., nucleic acid) testing. In others, the molecular alteration leading to the disease process lends itself to precise, reproducible testing by PCR or FISH, whereas still other diseases either have an unknown molecular mechanism or are otherwise not amenable to molecular testing. For each specific disease, the advantages and limitations of these methods must be considered before ordering studies that can give misleading or conflicting results. By the thoughtful use of these tests in concert, a more precise classification of disease entities and a better understanding of their pathogenesis can be achieved.

REFERENCES

1. Coons, A. H., Creech, H. J., and Jones R. N. Immunological properties of an antibody containing a fluorescent group. *Proc. Soc. Exp. Biol.* **47:**200–202, 1941.
2. Taylor, C. R., and Burns, J. The demonstration of plasma cells and other immunoglobulin containing cells in formalin-fixed, paraffin-embedded tissues using peroxidase labeled antibody. *J. Clin. Pathol.* **27:**14–20, 1974.
3. Kohler, G., and Milstein C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256:**495–497, 1975.
4. Huang, S.-N. Immunohistochemical demonstration of hepatitis B core and surface antigens in paraffin sections. *Lab. Invest.* **33:**88–95, 1975.
5. Shi, S. R., Key, M. E., and Kalra, K. L. Antigen retrieval in formalin-fixed, paraffin-embedded tissues: an enhancement method for immunohistochemical staining based on microwave oven heating of tissue sections. *J. Histochem. Cytochem.* **39:**741–748, 1991.
6. Dabbs, D. J. *Diagnostic Immunohistochemistry.* Churchill, New York, Livingstone, 2002.
7. Taylor, C. *Immunomicroscopy: A Diagnostic Tool for the Surgical Pathologist.* Major Problems in Pathology Vol. 19, 1994.
8. Hsi, E. D. A practical approach for evaluating new antibodies in the clinical immunohistochemistry laboratory. *Arch. Pathol. Lab. Med.* **125:**289–294, 2001.
9. Werner, M., Chott, A., Fabriano, A., and Battifora, H. Effect of formalin tissue fixation and processing on immunohistochemistry. *Am. J. Surg. Pathol.* **24:**1016–1019, 2000.
10. Shi, S. R., Cote, R. J., and Taylor, C. R. Antigen retrieval immunohistochemistry: past, present, and future. *J. Histochem. Cytochem.* **45:**327–343, 1997.
11. Shi, S. R., Cote, R. J., and Taylor, C. R. Antigen retrieval immunohistochemistry and molecular morphology in the year 2001. *Appl. Immunohistochem. Mol. Morphol*. **9:**107–116, 2001.
12. Huang, S.-N. Immunohistochemical demonstration of hepatitis B core and surface antigens in paraffin sections. *Lab. Invest.* **33:**88–95, 1975.
13. Rodney, M. T., Swanson, P., and Wick, M. R. Fixation and epitope retrieval in diagnostic immunohistochemistry: a concise review with practical considerations. *Appl. Immunohistochem.* **8:**228–235, 2000.
14. Cattoretti, G. and Suurmeijer, A. J. H. Antigen unmasking on formalin-fixed paraffin-embedded tissue using microwaves: a review. *Adv. Anat. Pathol.* **2:**2–9, 1995.
15. Cuevas, E. C., Bateman, A. C., Wilkins, B. S., et al. Microwave antigen retrieval in immunohistochemistry: a study of 80 antibodies. *J. Clin. Pathol.* **47:**448–452, 1994.
16. Taylor, C. R., Shi, S. R., Chaiwun, B., Young, L., Imam, A., and Cote, R. J. Strategies for improving the immunohistochemical staining of various intranuclear prognostic markers in formalin-paraffin sections: androgen receptor, estrogen receptor, progesteron receptor, p53, PCNA, and Ki-67 antigen revealed by antigen retrieval technique. *Hum. Pathol.* **25:**1107–1109, 1994.
17. Linden, M. D., Nathanson, S. D., and Zarbo, R. J. Evaluation of p53 antibody staining immunoreactivity in benign tumors and nonneoplastic tissues. *Appl. Immunohistochem.* **3:**232–238, 1995.
18. Hsu, S. M., Raine, L., and Fanger, H. Use of avidin–biotin peroxidase complex (ABC) in immunoperoxidase techniques: a comparison between ABC and unlabeled antibody (PAP) procedures. *J. Histochem. Cytochem.* **29:**577–580, 1981.
19. Bisgaard, K. and Pluzek, K. Use of polymer conjugates in immunohistochemistry: a comparative study of a traditional staining method to a staining method using polymer conjugates. *Pathol. Int.* **46(Suppl. 1):**577, 1996.
20. Vyberg, M. and Nielsen, S. Dextran: polymer conjugate two-step visualization system for immunohistochemistry. *Appl. Immunohistochem.* **6:**3–10, 1998.
21. Sabattini, E., Bisgaard, K., Ascani, S., et al. The Envision++ system: a new immunohistochemical method for diagnostic and research; critical comparison with APAAP, ChemMate, CSA, LABC, SABC techniques. *J. Clin. Pathol.* **51:**506–511, 1998.
22. Shi, S., Guo, J., Cote, R. J., et al. Sensitivity and detection efficiency of a novel two-step detection system (PowerVision) for immunohistochemistry *Appl. Immunohistochem. Mol. Morphol.* **7:**201–208, 1999.
23. Richter, T., Nahrig, J., Komminoth, P., Kowolik, J., and Werner, M. Protocol for ultrarapid immunostaining of frozen sections. *J. Clin. Pathol.* **52:**461–463. 1999.
24. Bobrow, M. N., Harris, T. D., Shaughnessy, K. J., and Litt, G. J. Catalyzed reporter deposition, a novel method of signal amplification; application to immunoassays. *J. Immunol. Methods.* **125:**279–289, 1989.
25. Adams, J. C. Biotin amplification of biotin and horseradish peroxidase signals in histochemical stains. *J. Histochem. Cytochem.* **40:** 457–1463, 1992.

26. Grogan, T. M. Automated immunohistochemical analysis. *Am. J. Clin. Pathol*. 98(**Suppl.** 1):S35–S38, 1992.
27. Le Neel, T., Moreau, A., Laboisse, C., and Truchaud, A. Comparative evaluation of automated systems in immunohistochemistry. *Clin. Chim. Acta.* **278:**185–192, 1998.
28. Seidal, T., Balaton, A., and Battifora, H. Interpretation and quantification of immunostains. *Am. J. Surg. Pathol.* **25:**1204–1207, 2001.
29. Van Wasielewski, R., Mengel, M., Wiese, B., Rüdiger, T., Müller-Hermelink, H. K., and Kreipe, H. Tissue array technology for testing interlaboratory and interobserver reproducibility of immunohistochemical estrogen receptor analysis in a large multicenter trial. *Am. J. Clin. Pathol.* **118:**675–682, 2002.
30. Cartun, R. W. Immunohistochemistry in infectious diseases. *J. Histotechnol.* **18:**195–202, 1995.
31. Cohen, J. Epstein–Barr virus infection. *N. Engl. J. Med.* **343:** 481–492, 2000.
32. Hsu, J. L., and Glaser, S. L. Epstein–Barr virus-associated malignancies: epidemiologic patterns and etiologic implications. *Crit. Rev. Oncol. Hematol.* **34:**27–53, 2000.
33. Herrmann, K., and Niedobitek, G. Epstein–Barr virus-associated carcinomas: facts and fiction. *J. Pathol.* **199:**140–145, 2003.
34. Swaminathan, S. Molecular biology of Epstein–Barr virus and Kaposi's sarcoma-associated herpesvirus. *Semin. Hematol.* **40:**107–115, 2003.
35. Gulley, M. L. Molecular diagnosis of Epstein–Barr virus-related diseases. *J. Mol. Diagn.* **3:**1–19, 2001.
36. Courville, P., Simon, F., Le Pessot, F., Tallet, Y., Debab, Y., and Metayer, J. Detection of HHV8 latent nuclear antigen by immunohistochemistry: a new tool for differentiating Kaposi's sarcoma from its mimics. *Ann. Pathol.* **22:**264–276, 2002.
37. Negri, G., Egarter-Vigl, E., Kasal, A., Romano, F., Haitel, A., and Mian, C. p16INK4a is a useful marker for the diagnosis of adenocarcinoma of the cervix uteri and its precursors: an immunohistochemical study with immunocytochemical correlations. *Am. J. Surg. Pathol.* **27:**187–193, 2003.
38. Riethdorf, L., Riethdorf, S., Lee, K. R., Cviko, A., Loning, T., and Crum, C. P. Human papillomaviruses, expression of p16, and early endocervical glandular neoplasia. *Hum. Pathol.* **33:**899–904, 2002.
39. Chan, J. K. Advances in immunohistochemistry: impact on surgical pathology practice. *Semin. Diagn. Pathol.* **17:**170–177, 2000.
40. Kaufmann, O., Fietze, E., and Dietel, M. Immunohistochemical diagnosis in cancer metastasis of unknown primary tumor. *Pathologie* **23:**183–197, 2002.
41. Chu, P., Wu, E., and Weiss, L. M. Cytokeratin 7 and cytokeratin 20 expression in epithelial neoplasms: a survey of 435 cases. *Mod. Pathol.* **13:**962–972, 2000.
42. Harris, N. L., Jaffe, E. S., Stein, H., et al. A revised European–American classification of lymphoid neoplasms: a proposal from the International Lymphoma Study Group. *Blood* **84:**1361–1392, 1994.
43. Harris, N. L., Jaffe, E. S., Diebold, J., et al. World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: report of the Clinical Advisory Committee meeting–Airlie House, Virginia, November 1997. *J. Clin. Oncol.* **17:**3835–3849, 1999.
44. Torlakovic, E., Torlakovic, G., Nguyen, P. L., Brunning, R. D., and Delabie, J. The value of anti-pax-5 immunostaining in routinely fixed and paraffin-embedded sections: a novel pan pre-B and B-cell marker. *Am. J. Surg. Pathol.* **26:**1343–1350, 2002.
45. Wang, T., Lasota, J., Hanau, C. A., and Miettinen, M. Bcl-2 oncoprotein is widespread in lymphoid tissue and lymphomas but its differential expression in benign versus malignant follicles and monocytoid B-cell proliferations is of diagnostic value. *APMIS* **103:**655–662, 1995.
46. Kurtin, P. J., Hobday, K. S., Ziesmer, S., and Caron, B. L. Demonstration of distinct antigenic profiles of small B-cell lymphoma by paraffin section immunohistochemistry. *Am. J. Clin. Pathol.* **112:**319–329, 1999.
47. Zukerberg, L. R., Yang, W. I., Arnold, A., and Harris, N. L. Cyclin D1 expression in non-Hodgkin's lymphomas. Detection by immunohistochemistry. *Am. J. Clin. Pathol.* **103:**756–760, 1995.
48. Xu, Y., McKenna, R. W., Molberg, K. H., and Kroft, S. H. Clinicopathologic analysis of CD10+ and CD10- diffuse large B-cell lymphomas. Identification of a high-risk subset with coexpression of CD10 and bcl-2. *Am. J. Clin. Pathol.* **116:**183–190, 2001.
49. Lossos, I. S., Jones, C. D., Warnke, R., et al. Expression of a single gene, bcl-6, strongly predicts survival in patients with diffuse large B-cell lymphoma. *Blood* **98:**945–951, 2001.
50. Marshall-Taylor, C. E., Cartun, R. W., Mandich, D., and DiGiuseppe, J. A. Immunohistochemical detection of immunoglobulin light chain expression in B-cell non-Hodgkin lymphomas using paraffin-fixed, paraffin-embeded tissues and a heat-induced epitope retrieval technique. *Appl. Immunohistochem. Mol. Morphol.* **10:**258–262, 2002.
51. Leong, A. S.-Y., Yin, H., and Hafajee, Z. Patterns of immunoglobulin staining in paraffin-embedded malignant lymphomas. *Appl. Immunohistochem. Mol. Morphol.* **10:**110–114, 2002.
52. Kinney, M. C. The role of morphologic features, phenotype, genotype and anatomic site in defining extranodal T-cell or NK-cell neoplasms. *Am. J. Clin. Pathol.* **111(1 Suppl 1):**S104–S118, 1999.
53. Ohno, T., Stribley, J. A., Wu, G., Hinrichs, S. H., Weisenburger, D. D., and Chan, W. C. Conality in nodular lymphocyte-predominant Hodgkin's disease. *N. Engl. J. Med*. **14:**459–465, 1997.
54. Stein, H., Marafioti, T., Foss, H. D., et al. Down-regulation of BOB.1/OBF.1 and Oct2 in classical Hodgkin disease but not in lymphocyte predominant Hodgkin disease correlates with immunoglobulin transcription. *Blood* **97:**496–501, 2001.
55. Vasef, M. A., Alsabeh, R., Madeiros, L. J., and Weiss, L. M. Immunophenotype of Reed–Sternberg and Hodgkin's cells in sequential specimens of Hodgkin's disease: a paraffin section immunohistochemical study using heat-induced epitope retrieval method. *Am. J. Clin. Pathol.* **108:**54–59, 1997.
56. Kelly, K. M., Womer, R. B., Sorensen, P. H., Xiong, Q. B., and Barr, F. G. Common and variant gene fusions predict distinct clinical phenotypes in rhabdomyosarcomas. *J. Clin. Oncol.* **15:**1831–1836, 1997.
57. Delattre, O., Zucman, J., Melot, T., et.al. The Ewing family of tumors— a subgroup of small round cell tumors defined by specific chimeric transcripts. *N. Engl. J. Med.* **331:**294–299, 1994.
58. Gerald, W. L., Ladanyi, M., de Alava, E., Cuatrecasas, M., Kushner, B. H., and LaQuaglia, M. P. Clinical pathologic and molecular spectrum of tumors associated with t(11;22)(p13;q12): desmoplastic small round-cell tumor and its variants. *J. Clin. Oncol.* **6:**3028–3036, 1998.
59. Adams, V., Hany, M. A., Schmidt, M., Hassam, S., Briner, J., and Niggli, F. K. Detection of t(11;22)(q24;q12) translocation breakpoint in paraffin embedded tissue of the Ewing's sarcoma family by nested reverse transcription-polymerase chain reaction. *Diagn. Mol. Pathol.* **5:**107–113, 1996.
60. Parham, D. M., Webber, B., Holt, H., William, W. K., and Maurer, H. Immunohistochemical studies of childhood rhabdomyosarcomas and related neoplasms. Result of an intergroup rhabdomyosarcoma study project. *Cancer* **67:**3072–3080, 1991.
61. Wang, N. P., Marx, J., McNutt, M. A., Rutledge, J. C., and Gown, A. M. Expression of myogenic regulatory proteins (myogenin and Myo D1) in small blue cell tumors of childhood. *Am. J. Pathol.* **147:**1799–1810, 1995.
62. Cessna, M. H., Zhou, H., Perkins, S. L., et al. Are myogenin and myoD1 expression specific for rhabdomyosarcoma? A study of 150 cases with emphasis on spindle cell mimics. *Am. J. Surg. Pathol.* **25:**1150–1157, 2001.
63. Perlman, E. J., Dickman, P. S., Askin, F. B., Grier, H. E., Miser, J. S., and Link, M. P. Ewing's sarcoma—routine diagnostic utilization of MIC2 analysis —a Pediatric Oncology Group/Children's Cancer Group Intergroup Study. *Hum. Pathol.* **25:**304–307, 1994.
64. Folpe, A. L., Hill, C. E., Parham, D. M., O'Shea, P. A., Weiss, S. W. Immunohistochemical detection of FLI-1 protein expression: a study of 132 round cell tumors with emphasis on CD99-positive mimics of Ewing's sarcoma/primitive neuroectodermal tumor. *Am. J. Surg. Pathol.* **25:**1–12, 2001.
65. Ordonez, N. G. Desmoplastic small round cell tumor II: an ultrastructural and immunohistochemical study with emphasis on new immunohistochemical markers. *Am. J. Surg.Pathol.* **22:**1314–1327, 1998.

66. Charles, A. K., Moore, I. E., and Berry, P. J. Immunohistochemical detection of Wilms' tumor gene WT1 in desmoplastic small round cell tumor. *Histopathology* **30:**312–314, 1997.
67. Taylor, C. R. and Cote, R. J. Immunohistochemical markers of prognostic value in surgical pathology. *Histol. Histopathol.* **12:**1039–1055, 1997.
68. Harris, C. C. and Hollstein, M. Clinical implications of the p53 tumor-suppressor gene. *N. Engl. J. Med.* **329:**1318–1327, 1993.
69. Harris, C. C. Structure and function of the p53 tumor suppressor gene: clues for rational cancer therapeutic strategies. *J. Natl. Cancer. Inst*. **88:**1442–1455, 1996.
70. Hsu, F. D., Nielsen, T. O., Alkushi, A., et al.Tissue microarrays are an effective quality assurance tool for diagnostic immunohistochemistry. *Mod. Pathol.* **15:**1374–1380, 2002.
71. Falini, B. and Mason, D. Y. Proteins encoded by genes involved in chromosomal alterations in lymphoma and leukemia: clinical value of their detection by immunocytochemistry. *Blood* **99:**409–426, 2002.
72. Campbell, R. J. and Pignatelli, M. Molecular histology in the study of solid tumours.*Mol. Pathol.* **55:**80–82, 2002.
73. Gown, A. M. Geneogenic immunohistochemistry: a new era in diagnostic immunohistochemistry. *Curr. Diagn. Pathol.* **8:**193–200, 2002.
74. Paik, S., Hazan, R., Fisher, E. R., et al. Pathologic findings from the National Surgical Adjuvant Breast and Bowel Project: prognostic significance of erbB-2 protein overexpression in primary breast cancer. *J. Clin. Oncol*. **8:**103–112, 1990.
75. Pauletti, G., Dandekar, S., Rong, H., et al. Assessment of methods for tissue-based detection of the HER-2/neu alteration in human breast cancer: a direct comparison of fluorescence in situ hybridization and immunohistochemistry. *J. Clin. Oncol.* **18:**3651–3664, 2000.
76. Jacobs, T. W., Gown, A. M., Yaziji, H., Barnes, M. J., and Schnitt, S. J. Specificity of HercepTest in determining HER-2/neu status of breast cancers using the United States Food and Drug Administration-approved scoring system. *J. Clin. Oncol.* **17:**1983–1987, 1999.
77. Wang, S., Saboorian, M. H., Frenkel, E., Hynan, L., Gokaslan, S. T., and Ashfaq, R. Laboratory assessment of the status of Her-2/neu protein and oncogene in breast cancer specimens: comparison of immunohistochemistry assay with fluorescence in situ hybridisation assays. *J. Clin. Pathol.* **53:**374–381, 2000.
78. Zarbo, R. J. and Hammond, M. E. Conference summary, Strategic Science symposium. Her-2/neu testing of breast cancer patients in clinical practice. *Arch. Pathol. Lab. Med.* **127:**549–553, 2003.
79. Fletcher, C. D., Berman, J. J., Corless, C., et al. Diagnosis of gastrointestinal stromal tumors: a consensus approach. *Hum. Pathol.* **33:**459–465, 2002.
80. Hirota, S., Isozaki, K., Moriyama, Y., et al. Gain-of-function mutation of c-kit in human gastrointestinal stromal tumors. *Science* **279:**577–580, 1998.
81. Dematteo, R. P., Heinrich, M. C., El-Rifai, W. M., and Demetri, G. Clinical management of gastrointestinal stromal tumors: before and after STI-571. *Hum. Pathol.* **33:**466–477, 2002.
82. Greenson, J. K. Gastrointestinal stromal tumors and other mesenchymal lesions of the gut. *Mod. Pathol.* **16:**366–375, 2003.
83. Smith, M. E. F. and Pignatelli, M. The molecular histology of neoplasia: the role of the cadherin/catenin complex. *Histopathology.* **31:**107–111, 1997.
84. Droufakou, S., Deshmane, V., Roylance, R., Hanby, A., Tomlison, I., and Hart, I. R. Multiple ways of silencing E-cadherin gene expression in lobular carcinoma of the breast. *Int. J. Cancer* **92:**404–408, 2001.
85. Chan, J. K. and Wong, C. S. Loss of E-cadherin is the fundamental defect in diffuse-type gastric carcinoma and infiltrating lobular carcinoma of the breast. *Adv. Anat. Pathol.* **8:**165–172, 2001.
86. Marra, G. and Boland, C. R. Hereditary nonpolyposis colorectal cancer: the syndrome, the genes, and historical perspectives. *J. Natl. Cancer Inst.* **87:**1114–1125, 1995.
87. Lindor, N. M., Burgart, L. J., Leontovich, O., et al. Immunohistochemistry versus microsatellite instability testing in phenotyping colorectal tumors. *J. Clin. Oncol*. **20:**1043–1048, 2002.
88. Rigau, V., Sebbagh, N., Olschwang, S. et al. Microsatellite instability in colorectal carcinoma. The comparison of immunohistochemistry and molecular biology suggests a role for hMLH6 immunostaining. *Arch. Pathol. Lab. Med.* **127:**694–700, 2003.
89. Stein, H., Mason, D. Y., Gerdes, J., et al. The expression of the Hodgkin's disease associated antigen Ki-1 in reactive and neoplastic lymphoid tissue: evidence that Reed–Sternberg cells and histiocytic malignancies are derived from activated lymphoid cells. *Blood.* **66:**848–858, 1985.
90. Morris, S. W., Kirstein, M. N., Valentine, M. B., et al. Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* **263:**1281–1284, 1994.
91. Shiota, M., Fujimoto, J., Takenaga, M., et al. Diagnosis of t(2;5)(p23;q35)-associated Ki-1 lymphoma with immunohistochemistry. *Blood* **84:**3648–3652, 1994.
92. Falini, B., Pulford, K., Pucciarini, A., et al. Lymphomas expressing ALK fusion protein(s) other than NPM–ALK. *Blood* **94:**3509–3515, 1999.
93. Sen, F., Vega, F., and Medeiros L.J. Molecular genetic methods in the diagnosis of hematologic neoplasms. *Semin. Diagn. Pathol.* **19:**72–93, 2002.
94. Chan, J. K., Miller, K. D., Munson, P., and Isaacson, P. G. Immunostaining for cyclin D1 and the diagnosis of mantle cell lymphoma: is there a reliable method? *Histopathology.* **34:**266–270, 1999.
95. Leong, A. S. Y., Lee, E. S., and Yin, H. Superheating antigen retrieval. *Appl. Immunohistochem. Mol. Morphol.* **10:**263–268, 2002.
96. Kodet, R., Mrhalova, M., Krskova, L., et al. Mantle cell lymphoma: improved diagnostics using a combined approach of immunohistochemistry and identification of t(11;14)(q13;q32) by polymerase chain reaction and fluorescence in situ hybridization. *Virchow's. Arch.* **442:**538–547, 2003.
97. Belaud-Rotureau, M. A., Parrens, M., Dubus, P., Garroste, J. C., de Mascarel, A., and Merlio, J. P. A comparative analysis of FISH, RT-PCR, PCR and immunohistochemistry for the diagnosis of mantle cell lymphomas. *Mod. Pathol.* **15:**517–525, 2002.

17 Laser Capture Microdissection

C. Robert Bagnell, Jr.

1. INTRODUCTION

Advances in our understanding of disease mechanisms have resulted in the need for single-cell analysis. Analytical technologies have become available to accommodate such interrogations. Typically, molecular diagnostic assays begin with a nucleic acid extraction procedure during which tissue architecture and cellular morphology is lost. Laser capture microdissection (LCM) is a technology that enables scientists to examine the processes of individual cells. Whether one is investigating a cell's internal messages or its proteins, isolating that particular cell(s) from a mixed cellular environment is the function of LCM (Fig. 1). This chapter briefly describes the LCM technique by reviewing the current instrumentation and answers some of the most frequently asked questions about LCM.

There is now a vast literature on LCM, which this chapter will not attempt to review. A well-organized listing of the primary papers as well as contemporary work can be found at the Arcturus website http://www.arctur.com. Conn is editor of perhaps the best compilation to date on LCM *(1)*. There is a new methods book on LCM, edited by Murray and Curran *(2)*. LCM was introduced by the National Institutes of Health investigators Liotta, Bonner, and Emmert-Buck in 1996 *(3)* and 1997 *(4)*. The first commercial instrument was produced by Arcturus Engineering, Inc. (Mountain View, CA) as a result of a Cooperative Research and Development Agreement with NIH. To date, there are four companies that produce LCM equipment.

2. LCM METHODS AND INSTRUMENTATION

Three methods exist that use lasers to collect tiny samples from heterogeneous biological specimens. The first method melts cell-sized spots of a thermoplastic film onto the specimen using an infrared (IR) laser. The plastic cools adhering to the specimen. When the film is lifted, the adhering specimen is also removed. This is the original LCM method created at NIH and first reported by Liotta et al. *(3,4)*. This method is marketed by Arcturus (http://www.arctur.com). The second method uses a special supporting membrane under the specimen. A pulsed ultraviolet (UV) laser cuts the membrane around the desired specimen that then either drops into a collection cap by gravity (Leica; http://www.leica-microsystems.com) or is catapulted into a collection cap by a defocused laser beam (P.A.L.M., http://www.palm-mikrolaser.com). The third method uses an IR laser and a special IR-absorbing plastic film on which the sample is placed. The laser severs the film by heat, thus isolating islands of specimen which remain when the bulk of the film is removed (Bio-Rad; http://www.bio-rad.com).

2.1. ARCTURUS The Arcturus Pix-Cell IIe is shown in Fig. 2. An automated version called the AutoPix is also available. The Arcturus LCM process is illustrated in Fig. 3. The specimen is placed on a glass slide with no cover slip. Critical to this process is the LCM cap. The cap, which is sized to fit into a 0.5-mL Eppendorf tube, is made of an optical-grade plastic with a thermoplastic film on the narrow end. The film is placed on the specimen, which is visualized using an inverted microscope equipped with a color charge-coupled device (CCD) camera and video monitor. The specimen is maneuvered utilizing the joy-stick stage positioner to place an area of interest under a target beam that appears on the monitor. The IR laser is activated by pressing a button. The laser passes through the cap from above and causes the thermoplastic film to soften and expand down into the tissue at the position of the target beam. The film adheres to the tissue. Each press of the button is called a "shot." One cap can contain approx 6000 shots. After all areas to be microdissected have been shot, the cap with adherent specimen is lifted away from the slide. The cap is then placed into a 0.5-mL Eppendorf tube for processing of the specimen.

2.2. LEICA AS LMD Figure 4 illustrates the Leica AS LMD. This system is based on the ability of a pulsed UV laser to cut a PEN (polyethylene–naphthalate) membrane covering a glass slide with the specimen on top. The slide is placed on the computer-controlled stage of the upright microscope, specimen side down. Fig. 5. illustrates the stage area. Areas of interest are visualized utilizing the computer-controlled stage and focus system and a color CCD camera. Specimen regions to be microdissected are outlined on the video image using the mouse cursor. Once this is done, the computer system automatically moves the stage to each location. The system guides the UV laser cutting by deflecting the laser beam through the

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

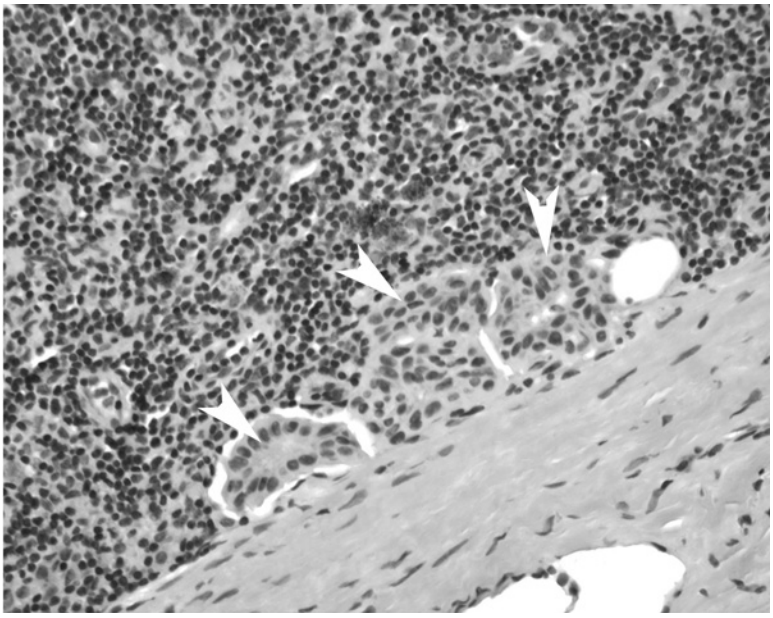


Fig. 1. Micrometastatic cancer cells in a lymph node (× 40 objective). Only the cells that are forming glandlike structures, indicated by the arrows, are of interest in analyzing the tumor cell genes or proteins.

Fig. 2. Arcturus PixCell IIe LCM system.

objective lens around the previously marked areas. Once cut, the PEN membrane with the specimen attached drops by gravity into the cap of a 0.5-mL Eppendorf tube held in place below the specimen by a special carrier. After microdissection, the cap and tube are removed from the carrier and are ready for further processing.

2.3. P.A.L.M. The P.A.L.M. system is illustrated in Fig. 6. The P.A.L.M. system also utilizes a pulsed UV laser for cutting. However, the specimen can be on a variety of substrates, including plain glass slides, slides with a PEN membrane, or culture dishes with a PEN membrane insert. The specimen is placed on the inverted microscope, specimen side up. A

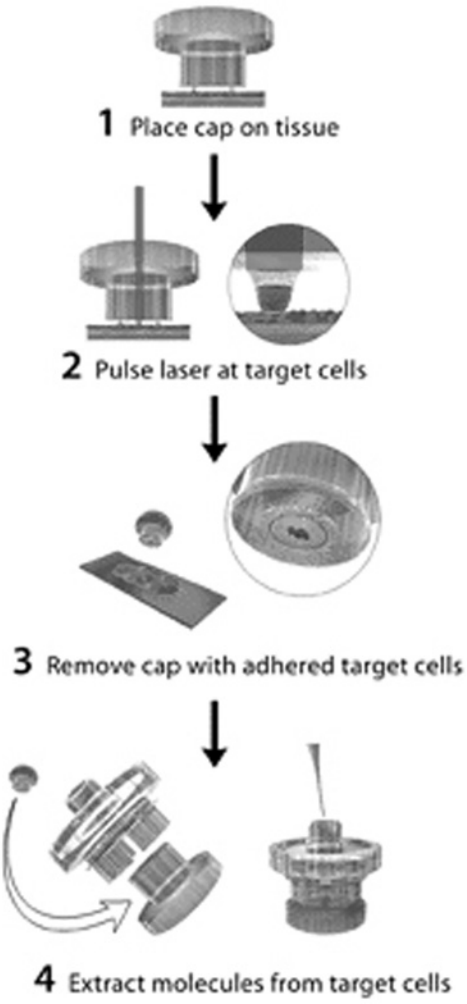


Fig. 3. Arcturus LCM process.

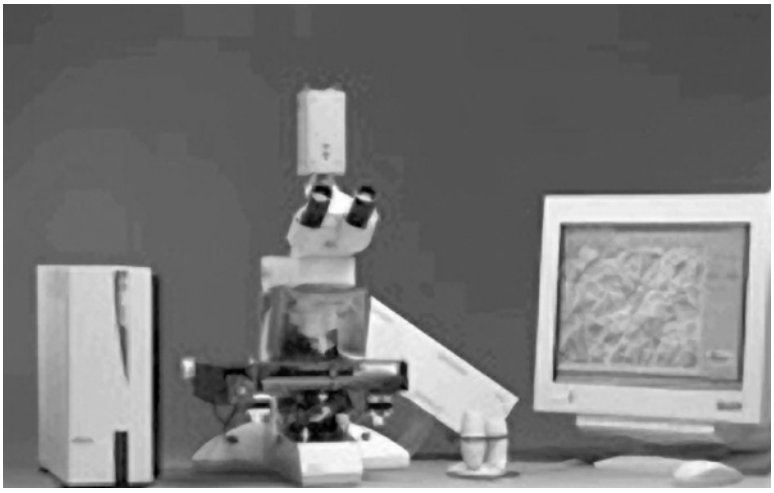


Fig. 4. Leica AS LMD system.

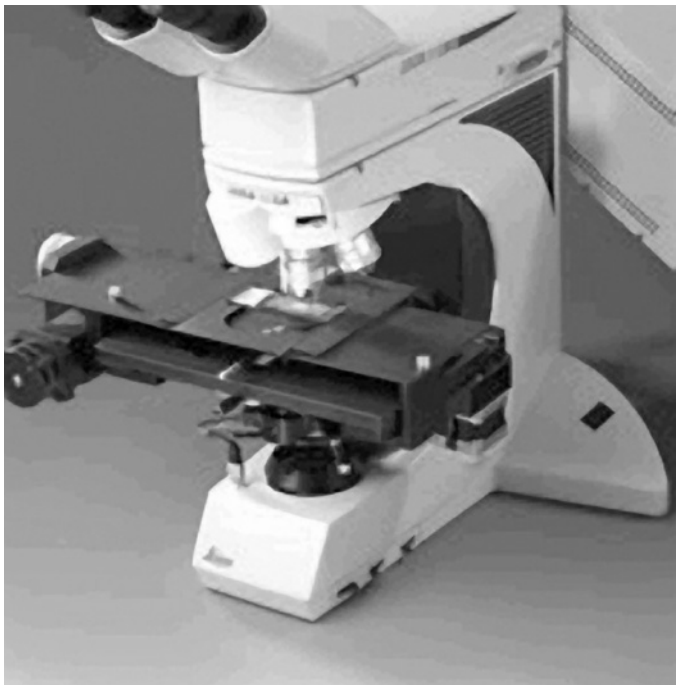


Fig. 5. Leica AS LMD stage.

computer controls the stage and microscope focus. A CCD camera presents an image of the specimen on the computer monitor. Marking areas for microdissection is carried out by outlining or "dotting" them in the video image using the mouse cursor. The system then moves to each location and cutting is done by projecting the laser through the objective lens while the computer moves the precision stage around a previously marked area. The specimen is then catapulted (removed by the force of light pressure in the defocused laser beam) into the cap of an Eppendorf tube held above the specimen in a special carrier. Specimens not on a PEN membrane can be microdissected by multiple catapult shots covering the desired area. The cap is then placed on a standard 0.5-mL Eppendorf tube for further processing.

2.4. BIO-RAD CLONIS The Bio-Rad CLONIS system is shown in Fig. 7. This system is based on the ability of an IR laser to cut a PEN membrane that is specially constructed to absorb the IR light energy, thus producing heat. Specimens must be placed on the special multilayered membrane whether on slides or in culture dishes. The sample is placed on the inverted microscope specimen side up. Areas for microdissection are selected utilizing the computer-controlled stage and CCD camera by marking them on the monitor utilizing the mouse cursor. After marking, the system moves to the marked areas and cuts the film by moving the stage over the fixed laser beam, which is projected through the objective lens. Samples are subsequently processed by either removing the membrane containing the unwanted sample or by removing the part containing the desired sample. Both procedures are done by hand, utilizing sharp forceps.

3. SYSTEM COMPARISONS (TABLE 1)

3.1. SPECIMEN CONDITIONS Of the four systems considered here, the Arcturus and Leica systems require that the samples be *dry*. Also, relative humidity is important for both of these systems. The Leica system requires at least 35% relative humidity (RH), otherwise static charges interfere with the gravity drop of the microdissected sample into the cap. The Arcturus system, on the other hand, requires a nonhumid atmosphere, otherwise the very dry sample will absorb room moisture and the film will not stick to the sample. The P.A.L.M. and Bio-Rad systems can microdissect both dry and wet specimens, including living specimens. The type of surface on which the sample is placed differs among the systems. Arcturus requires that the sample be on a surface that is less adherent to the specimen than the thermoplastic membrane; usually, plain, uncoated glass slides will work. The P.A.L.M., Leica, and Bio-Rad systems require that the specimen be placed on a special PEN membrane for laser cutting. The P.A.L.M. system can microdissect directly from a glass slide by the laser catapulting method.

Samples from a number of sources are suitable for microdissection. The main criterion is that the sample areas to be

Fig. 6. P.A.L.M. system.

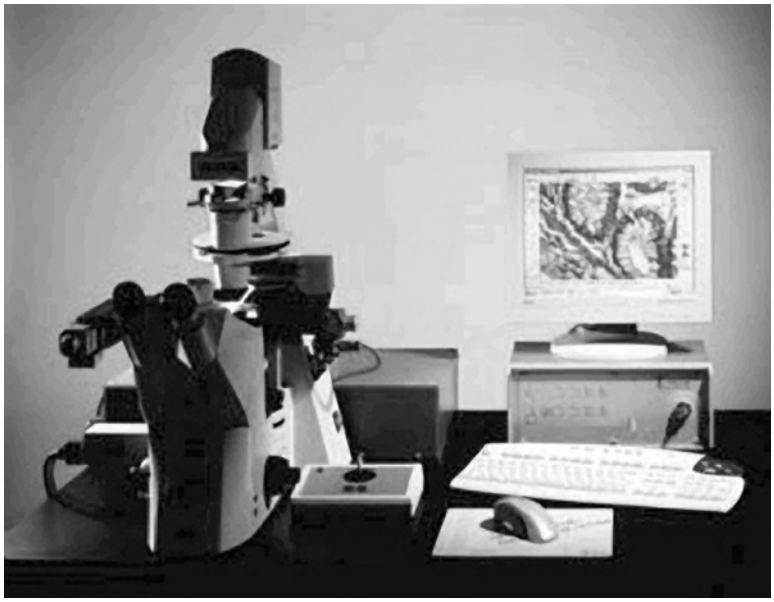


Fig. 7. Bio-Rad CLONIS.

microdissected must be visualized on a microscope. Frozen sections on glass slides, metaphase spreads on cover slips, aldehyde-fixed and paraffin-embedded tissue sections, cells on trans-well membranes, cytospins onto glass slides, and tissue cultures grown on specially coated slides or in dishes with special membrane inserts are some of the types of specimen that have been successfully microdissected utilizing the four systems described here. Several histological and immunological stains have been utilized to help visualize the specimen. Each investigator should do some testing to judge whether or not a particular stain has an affect on recovery of RNA, DNA, or proteins, or interferes with a polymerase chain reaction (PCR). Protocols for several histological stains can be found at the Arcturus website.

3.2. SMALLEST MICRODISSECTED AREA Laser spot size on the specimen sets the lower limit on the size of a microdissected area. Spot size is determined by the objective lens numerical aperture, the wavelength of the laser light, and the type and thickness of material the laser has to travel through to reach the sample. Typical laser spot size for the P.A.L.M. system is about 1 μm; for the Leica AS LMD, it about 2.5 μm, and for the Bio-Rad CLONIS, it about 20 μm. The Arcturus system has three spot sizes: 7.5, 15, and 30 μm. All of the systems can microdissect single cells, but for the Bio-Rad system, it is recommended that the cut area around a cell be about 200 μm away from the cell(s) of interest.

3.3. SPECIMEN VISUALIZATION Bright-field and fluorescence imaging are available on all systems. All but Arcturus system can also utilize phase-contrast and differential interference-contrast techniques. All of the systems offer good visualization of the specimen *prior* to microdissection. The Arcturus, Bio-Rad, and P.A.L.M. systems can visualize the sample *after*

Table 1
System Comparisons

Instrument	*Arcturus*	*P.A.L.M.*	*Leica*	*CLONIS*
Minimum dissectable area	7 µm	1 µm	1 µm	20 µm
Maximum dissectable area	30 µm	3 mm	3 mm	25 mm
Visualize dissected sample	Yes	Yes if cut /no if only catapulted	No	Yes
Visualize un-dissected sample	Yes	Yes	Yes	Yes
Thickness of dissected sample	5–10 µm	5–10 µm	5–10 µm	200 µm
Microdissect live cells	No	Yes	No	Yes
Specimen dehydration	Yes	No	Yes	No
Types of sample preparation	F, P, M	F, P, C	F, P	F, P, C
Specimen on special surface	No	Yes for cutting/no for only catapulting	Yes	Yes
Dissect multiple cell types simultaneously	No	Yes	Yes	Yes
Microscope type	I	I	U	I
Specimen up or down	Up	Up	Down	Up
Optical contrast methods	Bf, Fl	Bf, Ph, Fl	Bf, Ph, Fl	Bf, Ph, Fl
Computer-controlled stage	No	Yes	Yes	Yes
Laser wavelength (in nm)	840	336	336	840
Laser spot size (in µm)	7.5, 15, 30	1	2.5	20
Controlled laser ablation possible	No	Yes	Yes	Yes
Specimen collection method	TPF	LPC	G	MMR
Protocols	Yes	Yes	No	Yes

Abbreviations: F = frozen; P = paraffin; M = membrane filter; C = culture; I = inverted; U = upright; Bf = bright field; Fl = fluorescence; Ph = phase; TPF = thermoplastic film; LPC = laser pressure catapulting; G = gravity; MMR = manual membrane removal.

it is microdissected. In the case of the P.A.L.M., systems, this is possible as long as the sample has been cut from a PEN membrane. Samples that are microdissected by only catapulting directly from glass end up as flakes in the cap and show no recognizable morphology. In the case of the Bio-Rad system, visualizing the microdissected sample is possible if the microdissection is of the type that removes the unsampled material. For the Arcturus system, the cap with adherent specimen is placed down on a clean area of the slide and the microdissected samples can be seen quite clearly. All of the systems can visualize the area left after microdissection.

3.4. INSTRUMENT PROS AND CONS The Arcturus system is fast and simple to operate. In the Pix-Cell II series, no computer is required to do the microdissection. The attached computer can acquire images and perform other useful functions such as keeping track of the number of shots. It is easy to visualize the microdissected sample and to remove material not adhering to the membrane. For extremely heterogeneous samples in which only *one* component is desired, the Arcturus system is a very good system as long as the sample is *dry* The system should be placed in an air-conditioned facility because high RH can adversely affect adherence of the thermoplastic film to the specimen. The mechanical stage requires good manual dexterity to manipulate. Because the system utilizes a vacuum chuck to hold the sample on the microscope stage, materials for microdissection must be on glass slides and placed near the middle of the slide. The Arcturus system is the only system that cannot utilize the laser to ablate unwanted parts of the sample.

Leica AS LMD compares favorably with the Arcturus system. The computer-controlled system allows the user to choose many different areas for microdissection by marking them on the video monitor prior to any microdissection taking place. After marking, the system takes over and performs the microdissection. This is less tedious than the manual Arcturus method. Several different types of area can be marked and these will be microdissected into separate caps by the substage cap mechanism. The microdissected sample cannot be visualized directly. Samples must be placed on slides that are covered with a special PEN membrane. Because the system depends on gravity to collect the dissected sample into the collection cap, static electricity becomes a problem. Static charges on the cap, the PEN membrane, and other plastic materials can divert the dissected sample away from the cap. Therefore, the RH in the region of the cap must be around 35% or better to minimize static charges.

The P.A.L.M. system is the only system that can automatically microdissect both dry and wet samples directly into a collection cap. This is the most universally applicable system of those covered here. As with the Leica, several different types of area can be marked and these will be microdissected into separate caps by the above stage cap mechanism. Movement of the microdissected specimen into the cap is by a mechanism called laser catapulting—a partly understood mechanism in which light pressure from a pulse of the defocused laser is sufficient to move the specimen up into the collection cap *(5)*. Normally, samples should be placed on a PEN membrane either on slides or in culture dishes or on cover glasses if chromosomes are to be microdissected. Samples not on a PEN membrane, such as deparaffinized archival histological sections, can be microdissected by direct laser catapulting.

In this case, the microdissected sample cannot be visualized because it is highly fragmented.

Bio-Rad CLONIS is intended for microdissection of wet samples where the areas for microdissection are well separated from one another, although it can be used for dry samples as well. It is the only system that provides a mechanism for isolating islands of cells while still in culture by creating trenches in the membrane material that cells cannot cross. It can also easily ablate unwanted areas while the cells or tissue are still in culture. Tissues or cells must be placed on the special multilayer membrane. The thickness of the material that can be microdissected can be as great as 200 μm. Because the system uses an IR laser that must pass through the supporting medium, the spot size is quite large in the specimen plane (20–40 μm). This is not ideal for isolating single cells. The system requires that the cut membrane be removed or collected by hand utilizing fine forceps—a process requiring considerable manual dexterity.

4. QUESTIONS COMMON TO ALL LCM SYSTEMS

Preparation of the specimen prior to LCM is an area in which the scientist should exercise great caution and care. Judging from the literature, the level of knowledge is in great flux, much as electron microscopy was in the 1950s. There are some very good standard references such as ref.*6*. However, even there, no mention is made of the potential for hematoxylin staining to affect the consistency of DNA template amplification by PCR, as reported by Gulley et al. *(7)*. Specimen fixation definitely affects the quality of genetic material, with the consensus being that frozen specimens are best followed by alcohol fixation followed by formalin fixation *(8)*. Possible effects on the quality of genetic material caused by laser irradiation or laser-generated heat during microdissection are in question. Different LCM manufacturers make different claims. Liotta et al. in the original description of LCM, in which a pulsed nitrogen laser (operating in the UV-A range of 320–400 nm) melted a thermoplastic film onto the tissue, found that the brief 90°C temperature exposure of the tissue had no effect on subsequent PCR analysis or enzyme assays of the microdissected specimen *(3)*. Schütze et al. *(9)* claim no detrimental effects on RNA, DNA, or protein recovery from material catapulted using a UV-A laser.

The number of cells required for genetic analysis is generally in the range 300–500 cells for DNA and 500–1000 cells for RNA *(10)*. For proteins, 1000–5000 cells are required when using gel analysis *(11)* and as few as 100 cells using SELDI mass spectroscopy *(11)*.

ACKNOWLEDGMENTS

The author thanks the following companies for their generous help in providing materials and advice in preparing this manuscript and for permitting the use of images from their advertising literature:

Arcturus Engineering, Inc., 400 Longue Avenue, Mountain View, CA 94043–4019, USA, Tel: 888-446-7911; website: www.arctur.com

Bio-Rad Cell Science Division, Bio-Rad House, Maylands Avenue, Hemel Hempstead, HERTS HP27TD, UK; Tel: + 44 (0) 208 328 3228; website: www.cellscience.bio-rad.com

Leica Microsystems Inc., 111 Deer Lake Road, Deerfield IL 60015, USA; Tel: 800-248-0665; website: www.leica-microsystems.com

P.A.L.M. Microlaser Technologies AG, Am Neuland 9+12, 82347 Bernried, Germany; Tel: +49 (0) 8158-9971-0; E-mail:info@palm-microlaser.com website: www.palm-microlaser.com

NOTES

1. LCM refers to Laser Capture Microdissection and is specific to Arcturus. Laser Microdissection (LMD) is the more generic term. I have used LCM here for both.
2. The mmi Cellcut (www.molecular-machines.com) and Arcturus VERITAS appeared too late to be included in this chapter.
3. As of 2005, P.A.L.M. and CLONIS are marketed through Carl Zeiss (www.smt.zeiss.com).

REFERENCES

1. Conn, P. M. (ed.). *Laser Capture Microscopy and Microdissection*, Methods in Enzymology Vol. 365, Academic, San Diego, CA, 2002.
2. Murray, G. I. and Curran, S. *Laser Capture Microdissection: Methods and Protocols*. Methods in Molecular Biology Vol. 293. Humana, Totowa, NJ, 2005.
3. Emmert-Buck, M. R., Bonner, R. F., Smith, P. D., et al. Laser capture microdissection. *Science* **274:**998, 1996.
4. Bonner, R. F., Emmert-Buck, M. R., Cole, K., et al. Laser capture microdissection: molecular analysis of tissue. *Science* **278:** 1481, 1997.
5. Schüze, K. and Lahr, G. Identification of expressed genes by laser-mediated manipulation of single cells. *Nat. Biotechnol.* **16:**737, 1998.
6. Frost, A. R., Eltoum, I. E.,and Siegal, G. P. Nucleic acid amplification from individual cells. *Curr. Protocols Mol. Biol.* 25A.1.1–25A.1.24, 2001.
7. Burton, M. P., Schneider, B. G., Brown, R., Escamilla-Ponce, N., and Gulley, M. L. Comparison of histologic stains for use in PCR analysis of microdissected, paraffin-embedded tissues. *Biotechniques* **24:**86–92, 1998.
8. Simone, N. L., Bonner, R. F., Gillespie, J. W., Emmert-Buck, M. R., and Liotta, L. A. Laser-capture microdissection: opening the microscopic frontier to molecular analysis. *Trends Genet.***14 (7):**272–276, 1998.
9. Westphal, G., Burgemeister, R., Friedemann, G., et al. Noncontact laser catapulting: a basic procedure for functional genomics and proteomics, in *Laser Capture Microscopy and Microdissection,* Methods in Enzymology Vol. 365, Conn, P. M. (ed.), Academic, San Diego, CA, pp. 80–99, 2002.
10. Wittliff, J. L. and Erlander, M. G. Laser capture microdissection and its applications in geonomics and proteomics, in *Laser Capture Microscopy and Microdissection*, Methods in Enzymology Vol. 365, Conn, P. M. (ed.), Academic, San Diego, CA, pp. 12–25, 2002.
11. Craven, R. A. and Banks, R. E. Use of laser capture microdissection to selectively obtain distinct populations of cells for proteomic analysis, in *Laser Capture Microscopy and Microdissection,* Methods in Enzymology Vol. 365, Conn, P. M., ed., Academic, San Diego, CA, pp. 33–49, 2002.

V

Quality Assurance in the Molecular Diagnostics Laboratory

18 Framework for Quality Assurance in Molecular Diagnostics

MARLENE SABBATH-SOLITARE, SELWYN J. BAPTIST, AND TERESITA CUYEGKENG REDONDO

1. INTRODUCTION

A framework for ensuring laboratory quality was laid down by Congress in the Clinical Laboratory Improvement Act of 1967. Because of public concerns about the quality of clinical laboratory testing in the United States, Congress passed the Clinical Laboratory Improvement Amendments of 1988 (CLIA'88) setting forth uniform quality standards for all laboratories performing tests for health purposes on human specimens. Laboratories must register with the Department of Health and Human Services (HHS) to obtain and maintain a current CLIA certificate regardless of whether they receive payment from Medicare or Medicaid programs. The Centers for Medicare & Medicaid Services (CMS), formerly the Health Care Financing Administration (HCFA), is a federal agency within HHS and is charged with the implementation of the CLIA'88 regulations. CMS working with the Public Health Services (PHS), specifically the Centers for Disease Control and Prevention (CDC) and the Food and Drug Administration (FDA), developed standards for laboratory certification and set forth criteria for categorizing tests based on the level of complexity to perform the test. Tests, therefore, are categorized as waived, moderate complexity including the subcategory of provider-performed microscopy, or high complexity. Waived tests employ methods that are simple and are assumed to pose no risk or harm to the patient if the test is not performed correctly. Laboratories performing waived tests must follow manufacturers' instructions for test performance and are not subject to routine inspections. Tests categorized as moderate and high complexity must meet such CLIA'88 requirements as: (1) maintaining optimal patient specimen integrity and identification throughout the testing process, (2) specifying the responsibilities and qualification for personnel performing the test, (3) establishing and following written Quality Control procedures and (4) having a comprehensive Quality Assurance program in place. The laboratories performing moderate- and high-complexity testing must participate in a proficiency testing program for each specialty, analyte, or test for which they are certified and are subject to inspections by CMS or other private accrediting organizations. The major differences between moderate- and high-complexity testing are in the quality control and personnel standards. On February 29, 1992, these revised CLIA'88 regulations were published in the *Federal Register (1)* and it is recommended that each laboratory have a copy.

On January 24, 2003, almost 11 yr later, CMS and the CDC published a final rule in the *Federal Register (2)* revising the quality control and personnel standards for laboratory services and changing the consensus requirements for grading proficiency tests. The rule specifies a common set of quality control (QC) standards for both moderate and high complexity testing (now known as nonwaived testing) and reduces the frequency with which QC is to be done in most specialty and subspecialty areas. Under the new rules, agencies providing proficiency testing programs must grade a laboratory's sample in cases where there is 80%, not the current 90%, agreement among the participants or referee laboratories. This will ensure that more laboratories will be graded and ultimately improve the accuracy of laboratory testing. The CLIA rules now put the onus on the laboratory to review and validate nongraded samples, especially when the results do not agree with the intended response. Another change in the regulations is that directors of high-complexity testing are to be board certified unless they are eligible to be grandfathered under the phase-in requirements.

The CLIA'88 standards were designed to enhance patient safety while making it easier for laboratories to read, understand, and comply with the requirements. Subpart J (Patient Test Management), Subpart K (Quality Control), and Subpart P (Quality Assurance) were consolidated and reorganized into a new Subpart J (Facility Administration for Nonwaived Testing) and a new Subpart K (Quality Systems for Nonwaived Testing). Subpart J clarifies requirements for facility space, utilities, safety, transfusion services, as well as record and specimen retention. Subpart K pertains to the total testing process. The reorganization is written to parallel the flow of a patient specimen through the laboratory from the acquisition of a specimen

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

with the test request, to test performance and reporting of results. Each phase of the testing process now has its own standards. In this way, laboratories have a better chance to identify and prevent errors. This subpart of the rules also incorporates the requirements of quality assurance (QA) (renamed quality assessment to more clearly reflect the activities performed) under each appropriate section, namely: General Laboratory Systems, Preanalytic Systems, Analytic Systems, and Post-analytic Systems, to ensure that quality services are provided throughout the testing process. The essential component of the QA program is for laboratories to ensure continuous improvement of their performance and services through monitoring and evaluating the effectiveness of their policies and procedures for each phase of the testing system. They are to identify problems and take corrective action, to revise policies to prevent recurrences of problems, and to assess the adequacy and competency of the laboratory staff. Essential to the assessement activities is documentation.

There are quality system standards for 19 different nonwaived testing subspecialties that have been set forth in Subpart K of the final CLIA'88 regulations. No standards exist specifically for molecular testing, except for cytogenetics. Because laboratories that test human specimens must comply with CLIA'88 regulations, there are agencies reviewing the current regulations to determine which are applicable or should be amended to cover molecular testing. Until such time, molecular testing laboratories must follow good laboratory practice guidelines and participate in external quality assessment programs because they are considered in the nonwaived testing category. Both federal and state agencies (CLIA and New York State, Department of Health) and professional organizations (College of American Pathologists, American College of Medical Genetics and National Committee for Clinical Laboratory Standards) have available guidelines, recommendation, and checklists pertaining to quality control, quality assurance, proficiency testing, and personnel requirements. Many of the recommendations of the American College of Medical Genetics (ACMG) in the third edition of *Standards and Guidelines for Clinical Genetics Laboratories (3)* have been endorsed and adopted by the College of American Pathologists (CAP). Jointly, the ACMG and CAP in their proficiency testing program offer surveys that are designed to emphasize the preanalytic and postanalytic phases of molecular genetic testing and molecular oncology testing *(4).* The CAP also offers onsite laboratory inspection and a certification program. CAP has been given deemed status by CMS as a laboratory certifying agency under CLIA'88 and is also recognized by the Joint Commission on Accreditation of Healthcare Organization (JCAHO). As such, any laboratory providing molecular testing can seek certification through the CAP Laboratory Accreditation Program as a means to comply with CLIA'88 regulations. CLSI (National Committee for Clinical Laboratory Standards) is a voluntary organization that develops consensus recommendations for the standardization of test methodologies within the healthcare field. Its *Molecular Diagnostic Methods for Genetic Diseases; Approved Guidelines (5)* covers all phases of operation of a molecular genetics diagnostic laboratory, including: nomenclature, laboratory safety, specimen handling and testing, quality assurance and results reporting.

Even though components of a quality assurance program (preanalytical, analytical, and postanalytical) are well established and consistent among published guidelines, there are concerns that safeguards are not in place to ensure high-quality service in molecular testing laboratories. A report by McGovern *(6)* concluded that a number of molecular genetic testing laboratories had suboptimal quality assurance practices and suggested that both personnel qualifications and laboratory practice standards need improvement. This was not substantiated by Hofgartner and Tait *(7),* who inferred that the majority of molecular genetic laboratories were accredited by the CAP, therefore, rigorous quality standards to improve the quality of genetic testing exist.

A framework for a quality assurance program as promulgated by the new CLIA'88 regulations and as guidelines offered by professional organizations as the American College of Medical Genetics (ACMG), College of American Pathologists (CAP), and CLSI (Clinical and Laboratory Standards Institute, formerly NCCLS or National Committee for Clinical Laboratory Standards) is discussed in this chapter. Emphasis will be placed on how to apply quality assurance standards to laboratories to molecular diagnostic testing.

2. QUALITY ASSURANCE (QUALITY ASSESSMENT)

A quality assurance (QA) program that is established, implemented, and maintained by a laboratory can help ensure high-quality results are provided in clinically relevant turnaround times. Its major goal is to minimize laboratory errors by continual assessment and subsequent improvement of the services provided. Such a program is to cover all steps of the testing operation, including the preanalytical, analytical, and postanalytical processes, which must be continuously monitored and assessed. A successful QA program should identify and monitor key indicators, preferably of high-risk, high-volume activities, on a regular basis, to assess quality and to detect possible ways to improve processes and/or products. The document describing the policies and procedures of a QA program should contain the elements of why there is a need to assess, what is to be monitored, how and when this is to be done, and by whom. It should also include criteria to assure the competency of the laboratory personnel and provide a means of assessing laboratory activities through proficiency testing and peer review.

2.1. PREANALYTIC PHASE The preanalytical phase of a quality assurance program addresses activities that occur prior to testing the patient's specimen. This phase of the testing process has not received the attention it should. The final rules of the Federal Register *(2)* state that all laboratories are required to identify ways to monitor, assess, and when indicated, correct problems that might occur when tests are requested and during specimen collection and handling. Thus, a vital link in the quality assurance process is the proper identification of patients and their specimens, because a mistake in either of these areas might affect and subsequently influence test results. Each laboratory, therefore, should establish clearly defined criteria for monitoring the proper collection, labeling, preservation, transportation, and storage of a specimen to be tested as part of their QA program. Included in the program are the reasons for the rejection of an unacceptable specimen and

what corrective action is to be taken. A well-designed manual or electronic test request form should be prepared by the laboratory, so as to obtain relevant information required to provide accurate results and appropriate report interpretations. If results are entered into a recording or laboratory information system, the laboratory must ensure the information is correct and provide a system for timely correction of both clerical and analytical errors. There should be written criteria determining whether an informed consent is a requirement for specific molecular assays.

2.1.1. Test Requests Manual or electronic test request forms should be designed to allow for sufficient identification of the patient (name and address, date of birth, gender) and physician (name, address, and phone number), the test(s) requested, and pertinent clinical information. Date and time of collection as well as type of specimen collected should be included. For some assays, such as molecular genetic tests, the requisition might require racial/ethnicity data or a pedigree (e.g., for linkage analysis). Specific requirements for parentage/forensic identity testing can be found in detail by referring to the CAP Molecular Pathology Checklist and the American Association of Blood Banks Standards. Completed manual or electronic test requests must accompany the patient's specimen before a laboratory accession number is assigned. Multiple specimens from the same patient are each to have their own accession number.

One of the important considerations in molecular testing is to ensure that the clinical indications are appropriate for the test requested. This is the responsibility of the laboratory director or other members of the medical laboratory staff. If clarification of a clinical condition or any additional information regarding the test request is required, it is to be discussed with the referring physician. If there is a change in the test requested after consultation, it is to be documented and kept as part of the patient's file.

2.1.2. Specimen Collection and Handling The quality of test results is influenced to a high degree by the proper collection and handling of specimens. Ideally, laboratories should have designated reception areas to ensure that specimens are received and stored to minimize the chances of errors and or mix-ups. Procedures should be in place to include details of proper labeling of specimens, method of collection of specimens from all sources, specimen preservation (especially if processing is delayed), and their proper transportation to the laboratory. The protocols must be consistent with good laboratory practice and in compliance with all relevant safety codes as stipulated by the Occupational Safety and Health Administration (OSHA) and their institution *(3,4)*.

When received in the laboratory, a specimen should have been properly labeled with patient name, a unique identifying number (e.g., medical record), source of specimen, date and time collected. At this time, the accession process begins where the condition of the specimen (sample volume sufficient for analysis, is sample clotted, is container intact), the request form, and the referral data are reviewed for completeness. The date and time of receipt and the laboratory accession number are recorded onto a label firmly attached to each specimen container. If there is insufficient information for the laboratory to uniquely identify the specimen, CLSI recommends that specimens be rejected. However, the sample still could be tested if the demographics are incomplete, provided the information will be obtained by fax or phone in a timely manner. It is advisable that the action taken to obtain required data be documented by laboratory personnel and placed in the patient's file.

In situations where molecular testing laboratories are a part of a hospital-based department or reference laboratory, an accession number specific for molecular tests could be computer generated *(8)*. This molecular accession number and patient name is carried throughout the testing process and helps to reduce transcription errors. Only authorized laboratory personnel will have access to a password-protected computer system and to accession patient information. Folders can be created for each patient, at this time, to include a worksheet specific for the test requested, billing information, and a file card with patient name/accession number/test results/date finalized. These can function as a backup if the computer system should malfunction. The final report and supporting test data will eventually be included in the folder. A log of all accessioned patient specimens that were stored must be maintained to allow prompt retrieval for repeat or additional testing *(4)*.

The proper collection of the patient sample will depend on the clinical diagnosis and the test requested. Molecular analysis of neoplasia, infectious diseases, or genetic disorders require specimens from tumor tissue, body fluids, or any nucleated cells. The use of amplification techniques now renders extremely small samples, such as buccal scrapings, urine specimens, dried blood spots, and thin slices of paraffin-embedded tissue, adequate for testing. Laboratories must have available written requirements for the proper collection of patient specimens. For example, molecular testing of blood and bone marrow samples are to be collected in nuclease-free EDTA or ACD (acid citrate dextrose) tubes, not sodium heparin, to inhibit clotting. Heparin reportedly acts as an inhibitor in amplification assays *(9)*. If a sample is received in a heparinized tube, it is recommended that the white blood cells be separated and washed with a physiological buffer before the nucleic acid extraction. This is especially true for RNA analysis, as heparinized plasma has not proven to be suited for quantitative reverse transcription–polymerase chain reaction (RT-PCR) testing, as exemplified by poor quality results *(10)*. Fresh tissue samples for molecular testing should be frozen soon after collection to prevent DNA degradation and sent to the molecular laboratory immediately. If not frozen, it can be placed in a holding media and stored at 4°C for a maximum of 48 h. Tissues that have been paraffin-embedded or specimens preserved in ethanol are stable at room temperature, with minimal DNA degradation. However, tissues fixed in Zenker's, B5, or Bouin's produce extensive DNA damage, making their use inappropriate for molecular studies.

Blood and bone marrow specimens for DNA analysis, if collected properly, can be stored at 4°C for 3–5 days and still yield high-quality DNA for any amplification assay or Southern blot analysis. Farkas et al. *(11)* evaluated the quality and stability of DNA within clinical specimens such as blood (white blood cells were isolated before storing) and solid tissue (placenta) when stored at several different temperatures for various lengths of time. It was concluded that leukocytes and solid tissues kept at 4°C or frozen are best for short- to intermediate-term storage. In

our laboratory, all blood and body fluid samples are refrigerated upon arrival (usually up to 3 d), whereas tissue is immediately stored at –70°C, until ready for processing.

Specimens for RNA analysis require specific collection and handling conditions. One of the common causes of the failure to obtain results is due to insufficient stabilization of RNA in the specimen prior to dispatch to the laboratory. If chaotropic agents, such as guanidium isothiocyanate, are added immediately after collection, the specimen, under these conditions, can be stored for 7 d at ambient temperature before testing. These agents are known to remove proteins and denaturing RNase, which would otherwise degrade RNA. If these agents are not available, consider freezing the specimen to keep the RNA stable. It has been demonstrated that blood samples requiring RNA testing can be stored at 4°C for up to 48 h if the plasma is separated *(12)*.

2.1.3. Rejection of Specimens Specimen rejection may be justified if there is improper handling or a delay in transportation to the laboratory. Unacceptable specimens for testing would be ones that are received thawed in the laboratory when they should have been kept frozen, collected in the wrong anticoagulant, clotted, contaminated, or contain insufficient quantity. However, sub-optimal collection or storage conditions of patient samples should be reviewed on a case-by-case basis. For example, small, improperly handled specimens may be suitable for amplification assays, yielding interpretable molecular results, but not suitable for Southern blotting where large quantities of high-quality DNA are required. In general, it is left to the discretion of the laboratory staff and director to determine whether or not the specimen should be rejected.

If a specimen is rejected, the laboratory personnel must document the occurrence. The documentation should include the date, time, and condition the specimen was received, notification of the referring physician, and whether another specimen was obtainable. This record is kept in the patient file and documented as part of the QA records.

2.1.4. Informed Consent For some molecular genetic tests, informed consent might be required by federal, state, or local laws before initiating the test. This ensures that the patient voluntarily agrees to testing and has some understanding of the reasons for this test. The level of consent depends on whether the genetic test is used for predictive or diagnostic purposes. The laboratory can be of assistance in determining the appropriate level of informed consent and refer to established guidelines *(4)*. The New York State Department of Health recommends consent from all patients who are having a genetic test; however, the referring physician could sign the test requisition indicating that she/he conveyed the required information to the patient *(13)*.

2.2. ANALYTICAL PHASE In the analytical phase, laboratories performing nonwaived testing must meet applicable requirements as set forth in the CLIA'88 regulations and established by CAP's Laboratory Accreditation Program. Such requirements include the availability of a procedure manual, specimen storage, criteria to obtain accurate and reliable test results, and a system to maintain patient records *(2)*. Particular attention will be given to the following: nucleic acid extraction, contamination affecting amplification assays, use of controls, validation of tests, maintenance of equipment, documentation of competency of personnel, and the laboratory's participation in internal and external assessment programs.

2.2.1. Procedure Manual The procedure manual must contain sufficient details of all laboratory assays written so that qualified laboratory personnel can perform them consistently and accurately. The procedure is to include the principle of the test, pertinent clinical significance, specimen requirements, including collection, storage, preservation, and transportation to laboratory, criteria for specimen acceptability and rejection, the reagents needed, the steps to perform the test, quality control measures, verification ranges, interpretation of results, and references and pretinent notes. The style and format of the procedure manual, noted in the CAP Molecular Pathology Checklist, is at the discretion of the Laboratory Director; however, CLSI provides guidelines in their G6P42-A4 publication *(14)*. Electronic procedure manuals are acceptable as long as they are available to all personnel. They are subject to the same controls as the paper version in that they are to be reviewed annually by the laboratory director (or designee). All new procedures or revisions of existing ones are also to be reviewed by the laboratory director (or designee) prior to implementation. If a procedure is discontinued, a copy, must be maintained for a minimum of 2 yr. The initial date of use and retirement date must be clearly indicated on the procedure.

2.2.2. Nucleic Acid Extraction and Specimen Storage The extraction of nucleic acid, DNA, or RNA, is a crucial part of sample handling. Any errors that occur cannot be rectified at later stages of the testing process. Considerable attention should be given by well-trained technologists to adhering to the procedure at this stage of the assay. Methods that rapidly isolate high-quality DNA suitable for analysis are widely available. Kits exist that provide all the components and procedures necessary for isolating DNA from whole blood, bone marrow samples, cultured cells, and so forth *(15)*. However, automated instruments that extract nucleic acid limit the hands on part of the extraction procedure, thereby reducing the chances that an error will occur during the extraction process.

Once DNA or RNA are extracted, the conditions under which these nucleic acids are stored to prevent any significant loss in quality or quantity is important. This is especially important if they are to be used in other molecular assays or retested. The storage conditions for DNA and RNA differ. DNA can be stored for years if kept at 4°C in a buffered solution, such as Tris-EDTA. Diluting and storing in water will easily degrade DNA after a few weeks and it no longer will be useful for any molecular assay. RNA suspended in water must be frozen at –70°C soon after extraction and can be stored long term as such. Freezing and thawing RNA more than three times affects its stability and recovery *(10)*.

All nucleic acid samples must carry the laboratory's accession number, including the date of preparation. This unique identifier would assist in tracing the sample through the testing procedure. There must be a schedule for retaining nucleic acid specimens as stated in the CAP Molecular Pathology Checklist.

2.2.3. Contamination A major concern for any diagnostic molecular laboratory performing nucleic acid amplification methods is the occurrence of false-positive results due to

sample-to-sample contamination of nucleic acid during testing or the "carry-over" of DNA from a previous amplification of the same target. Poor laboratory technique and lack of attention to detail by technologists create situations in which amplified products contaminate areas of the laboratory. To minimize the occurrence of false-positive results, laboratories should implement specific practices and procedures including universal standards. An immediate plan of action when a contamination problem is observed would be to cease all testing, discard potentially contaminated reagents and supplies, and appropriately clean work surfaces and equipment. Testing would only resume if new aliquots or lot numbers of reagents, sterilized glassware, pipette tips, filtered buffers, etc., are available for use. The source of the contamination and the process used in its elemination must be documented and become part of the laboratory's QA program. Avoiding or minimizing laboratory contamination requires an appropriate laboratory design and good laboratory practices.

2.2.3.1. Laboratory design Each laboratory should develop its own unidirectional workflow to avoid the possibility of amplicon contamination. Ideally, a laboratory performing amplification assays should be divided into three separate work areas: (1) a reagent preparation area or room, (2) an area for specimen preparation, and (3) an area for amplification and detection *(16)*.

The reagent preparation area (area 1) must be kept very clean and away from any amplified products or patient specimens. The specimen processing area (area 2) should be located as far away from the amplification and detection area (area 3). This separation reduces the possibility of aerosols from extracted specimens to contaminate amplified products leading to false-positive results. Personnel traveling from the specimen preparation to the reagent preparation or the amplification and detection areas must always change their laboratory coats and gloves. Dedicated laboratory coats should be available when going in and out of each area. Laboratory coats or gloves worn in area 3 must never be worn in areas 1 or 2. Careless adherence to these recommendations can lead to carryover contamination from amplified products. However, with the introduction of commercially licensed tests and newer methodologies, some of the above requirements may be reduced.

Prior to the preparation of a specimen for amplification, bench tops and pipettes are to be wiped down with 10% sodium hypochlorite (bleach) and rinsed with 70% alcohol and water to remove biohazardous agents and extraneous nucleic acids. Contaminating amplicons remaining on workbenches can be destroyed by ultraviolet (UV) irradiation from germicidal bulbs through the creation of thymine dimers resulting in nucleic acid unsuitable for analysis. The use of Class II Biological safety cabinets with UV bulbs and HEPA-filtered air or bench top dead-air boxes with UV light attachments provide a clean dust-free worktop. These units are ideal for areas where specimens are prepared for extraction and where reagents are prepared. The UV light should be turned on at least one-half hour before manipulating any nucleic acid at the workbench and turned off before placing hands in working area.

It is recommended that there be a dedicated set of pipettes with plugged (aerosol-barrier) tips or positive-displacement tips for each of the three areas. The exclusive use of pipettes and tips will eliminate cross-contamination of samples. Prior to use pipette tips can be autoclaved, while pipette barrels, test tube racks, and mini centrifuges wiped with 70% ethanol. They all can remain under the hood when the UV light is on. Equipment such as gel apparatus and combs, centrifuges, microtome blades for cutting paraffin-embedded tissue, etc., should also be cleaned with 70% ethanol to minimize cross contamination.

2.2.3.2. Laboratory practices Reagents used in amplification assays once prepared should be divided into aliquots, and stores in an are separate from where specimens are prepared or amplified. Aliquoting minimizes the number of repeated sampling from the same test tube and reduces the potential for cross-contamination. Lot numbers and stored quantities of reagents should be recorded so that if carry-over does occur the source can be easily identified. All components of an amplification assay, such as dNTPs, primers, buffers, Taq polymerase, and RNAse/DNAse-free water, should be added to the reaction tube in the reagent preparation area and securely capped. Each tube should be capped after the addition of the sample, before proceeding to the next tube. It is recommended that the tubes be subject to a quick centrifugation before uncapping to prevent aerosolization.

Carry-over of nonspecific PCR products can be prevented by chemical modification of amplified fragments. For example, inactivation of an amplified product can be done by substituting dUTP for dTTP, thereby generating uracil-containing fragments in the reaction mixture *(17)*. The bacterial enzyme uracil-*N*-glycosylase (UNG) is added to the reaction and incubated prior to amplification. All uracil-containing DNA carried over from previous PCR reactions are enzymatically destroyed. Another example that is a major problem in isolating high-quality RNA is contamination by ribonuclease (RNAse). Sources of these highly resistant RNA-degrading enzymes are glassware and the hands of laboratory personnel. The treatment of glassware with an RNAse inhibitor, such as diethylpyrocarbonate (DEPC), followed by autoclaving or baking in a 250°C oven for 4 h will inactivate RNAse. Sterile disposable pipette tips and Eppendorf tubes are reported to be RNAse free, but autoclaving again will further reduce the chances of contamination.

2.2.4. Controls A variety of controls must be included for each test system to ensure that the results obtained are valid. The selection, number, and preparation of controls for procedures involving nucleic acid extraction, restriction enzyme digestion, electrophoresis, etc., are determined by the laboratory. Controls that must be included for each assay system are negative, positive, sensitivity controls and a molecular weight marker.

A negative control, meaning a reagent control to which no template has been added, must be included in every amplification assay. This negative control assesses the quality of the reagents by detecting any contaminant or increased background signal due to a few molecules of contaminating sequences. A positive control, which also contains all reagents required for amplification, provides proof that the assay works and specific evidence of amplification for each mutation or genotype included in the test system. In a Southern blot assay, a gene rearrangement (bands other than germline) must be observed for each restriction enzyme digest. These tests need to be

repeated. In RT-RNA assays, a "no reverse transcriptase" control is added each time to the test. If a positive result is observed when reverse transcriptase is omitted from a reaction tube, it indicates that either the reagents or the sample is contaminated. Molecular weight markers used with gel electrophoresis must span the range of expected bands for that specific primer set or probe locus.

Each molecular assay should also incorporate a sensitivity or analytical control to determine the lowest level of an analyte that can be detected by that specific methodology. This sensitivity control is especially important in Southern blot and amplification assays. For the Southern blot technique the sensitivity control corresponds to above 5% of the tumor cell population and is to be included with each electrophoresis run. For amplification assays, a known positive sample for that primer set should be serially diluted to the lowest limit for detection of a monoclonal population and included with each analysis.

To determine whether the nucleic acid is amplifiable, a control using primers directed toward a second target within the DNA template is recommended. A positive result indicates that the template is amplifiable and no reaction inhibitors are present. In some instances, these primers can be incorporated into an assay (multiplexing), saving both time and reagents.

Positive control specimens may not always be readily available for use in various assays. There are a number of sources in which such specimens can be obtained. The American Type Culture Collection (ATCC) and other commercial sources offer many tumor cell lines, DNA, and viruses for purchases. See these websites for details: http://ccr.coriell.org/nigms and www.atcc.org. Other sources for obtaining controls are through interlaboratory exchanges, proficiency testing samples, or patient samples that have been tested in parallel with alternative technology *(14)*. There is a need to establish a disease consortia so that the samples for controls (especially positive ones) can be obtained for testing. The Association for Molecular Pathology (AMP) through its website (www.amp.org) has proven to be a valuable source for technical information and education in the applications of molecular diagnostics and a media for exchange of methodologies and ideas.

2.2.5. Test validation Test validation should be conducted before a new test is introduced for clinical diagnosis. Each laboratory must document that they have validated the tests they are offering. Validation ensures that the test meets acceptable performance standards and is appropriate for the population for whom it is intended. Because requirements for the proper in-house validation of a test are not always clearly defined, the Association for Molecular Pathology (AMP) provides recommendations for in-house development of molecular assays *(18)* and the College of American Pathologists Molecular Pathology checklist refers to standards for test validation. Under CLIA'88 each laboratory that introduces an FDA-cleared or -approved test must demonstrate that it can obtain performance specifications, such as accuracy and precision, comparable to those established by the manufacturer. The laboratory must verify that the manufacturer's normal values are appropriate for their patient population. Assays developed in-house often use reagents purchased under the analyte-specific reagent (ASR) rule. Before introducing an ASR-based test, the laboratory must establish the following performance characteristics prior to reporting patient test results *(2)*.

- accuracy (verified by using split patient samples with another laboratory already performing the test and by using a recognized national agency that provides proficiency test sample, i.e., CAP)
- precision (can be established by using different levels of controls to determine within-run and between run and total imprecision)
- reportable range of the test results
- reference (normal) ranges
- analytic sensitivity (or lowest detectable limit)
- analytic specificity (the ability of the test to only detect the measurable quantity)
- clinical sensitivity (the percentage of positive tests when the clinical disorder is present)
- clinical specificity (the percentage of negative tests when the clinical disorder is absent)
- the positive predictive value (the likelihood that the clinical disorder is present when the test is positive) and
- the negative predictive value (the likelihood that the clinical disorder is absent when the test is negative).

The actual experiment performed to establish analytical and clinical performance characteristics vary and are dependent on whether it is a qualitative, semiquantitative, or quantitative type of test. Guidance for these experiments/evaluations can be obtained from some of the NCCLS documents (ww.nccls.org) or from the ACMG Standards and Guidelines specifically for molecular technique *(3)*.

Once the test is validated, it is recommended that some general quality assurance parameters be applied. These would include monitoring the effect the test results has no patient care, determining how useful the test results are for clinical management, and determining whether the cost of doing the test justifies the result in saving through more rapid and accurate diagnosis.

2.2.6. Maintenance of Equipment An organized system for maintaining and monitoring all equipment and instruments must be established by the laboratory to check the critical operating characteristics so as to provide consistent and reliable test results. Periodic maintenance and function checks must be done in accordance with the manufacturer's specifications. They are to be documented and kept at the workbench along with any service or repair records. For an instrument without specified function checks it is the laboratory's responsibility to establish maintenance and function checks. This system should be in compliance with CLIA'88 standards [see Standard: Maintenance and function checks 493.1254 in *Federal Register (2)*] and as stated in the CAP Molecular Pathology Checklist *(4)*.

2.2.7. Competency of Personnel The competency of technologists who perform clinical laboratory tests is to be assessed annually. The Laboratory Director, or designee, selects the process to be measured, indicates the reason for the selection, and describes the details of the task. Evaluation could either entail observing a specific aspect of a laboratory procedure, the operation of an instrument, or important laboratory safety procedures. It could also involve discussions with the technologist regarding a clinical procedure to assess their level of understanding or a review of issues pertaining to the control

and prevention of contamination when performing amplification assays in the laboratory. A specific example would be monitoring how a technologist in a molecular laboratory extracts DNA from peripheral blood using a specific manufacturer's kit.

The text of the evaluation should indicate the methodology used and the criteria anticipated for a satisfactory performance. If the competency is less than 100%, an explanation of the problem is to be provided, including any corrective or additional training measures to be implemented within a realistic timeframe. The evaluation is to be documented as well as signed and dated by the evaluator.

2.2.8. Proficiency Testing and Accreditation The goal of proficiency testing (PT) is to evaluate a laboratory's performance compared to other laboratories (its peers) participating in the PT program. It is a process for assessing laboratory performance by which multiple laboratories analyze similar samples that are collected and evaluated by an outside entity *(19)*. All CLIA laboratories performing nonwaived testing must participate in organized PT programs when available. When one does not exist for the test, there should be an external or internal PT program established by the laboratory. Applicable external procedures include the splitting of samples for analyses with a reference or other laboratory certified to perform the test or participating in an ungraded PT program, as organized by the Association for Molecular Pathology. For the internal program, some examples are to retest specimens previously assayed by laboratory personnel using pseudonym identifiers (blind study) or to perform clinical validation by chart review. In the blind-study PT method, the Laboratory Director selects the specimen, evaluates the results, and documents the outcome indicating that it is part of the internal PT program.

The College of American Pathologists requires laboratories to participate in proficiency testing or a CAP-approved program of graded interlaboratory comparison testing as part of its accreditation process. The laboratory must enroll in programs with analytes matching those tests that are performed for clinical testing. For molecular diagnostics, ACMG/CAP provides surveys for Molecular Genetics and Molecular Pathology. The laboratories are sent four to six specimens twice a year for each survey. The survey samples must be integrated within the routine laboratory workload using the same method as for patient samples and rotated among all laboratory personnel. A detailed questionnaire accompanies the samples and is used to record results and interpretation. The results are anonymously reviewed and analyzed. The final graded report, when issued to each of the participants, is an indication of the laboratory's performance compared to other laboratories performing the same procedures using identical or similar methods.

Unacceptable responses might be the result of improper handling of the PT sample, incorrect recording of results, or the incompetency of the testing personnel. Therefore, if an unacceptable result is received by the laboratory, the findings must be evaluated and corrective action taken. This action might entail personnel training and reviewing of test procedures. Both CLIA and CAP require documentation of action of unacceptable results.

Through the PT process, participants can identify procedural problems and take corrective action before patient results are affected. It can be used to document personnel competency and identify problems that might be resolved with additional training and monitoring. Proficiency testing provides continuing medical education for the laboratory.

2.3. POSTANALYTICAL PHASE The postanalytic phase issues involve assuring the accuracy and reliability of test results, reports designed to provide patient laboratory data effectively, and maintain patient confidentiality. The final report should provide an appropriate summary of the methods, the objective findings, and a clinical interpretation. Laboratory test results must be readily available to the laboratory and released to the requesting physician in a timely manner.

2.3.1. Laboratory Test Reports The laboratory report should include the name and address of the patient, a unique patient identifier, name and location of the laboratory, the date the test was performed, and description of the test methodology. The report should also include the source of the specimen and any information regarding the condition of the specimens if it did not meet the laboratory's criteria of acceptance. All laboratory reports must be clear, concise, accurate, and fully interpretative. The results should be brief and unambiguous. If they are complex, results may be presented in table format rather than text form. A statement interpreting the data, including clinical implications, follow-up recommendations, and the limits of the assay, must be written so that it is understandable to a nonmedical professional. The final report is to be signed by the laboratory director or other authorized individual and indicate the date it was finalized.

Molecular testing laboratories must include as part of the report the federally required clarifying statement for tests using analyte-specific reagents. The mandatory language is "This test was developed and its performance characteristics determined by (laboratory name). It has not been cleared or approved by the U.S. Food and Drug Administration." The CAP recommends the addition of "The FDA has determined that such clearance or approval is not necessary. This test is used for clinical purposes. It should not be regarded as investigational or for research. This laboratory is certified under the Clinical Laboratory Improvement Amendments of 1988 (CLIA'88) as qualified to perform high complexity clinical laboratory testing" *(4)*. It is our suggestion that FDA approved tests also be recognized as such in final reports.

A copy of the final report, including all supporting test data, are to be kept in the patient's file. Requests for copies to be sent to other health care providers are to documented by a written request from the patient. The reports, when finalized, are often distributed to the requesting physician via standard mail, or fax, or by hand.

Any questionable findings must be resolved by additional analysis (reprobing of blot or digestion of another sample, etc.) before reporting the results to the referring physician. If there is a lack of correlation of molecular results to other laboratory findings, such as histopathologic diagnosis, it is recommended that these cases be investigated to resolve the discrepancies. It could be just a problem with tissue sampling. Guidelines standardizing the reporting of results, interpretations, when to repeat a test, recommendations as to secondary testing, and correlation of results with other clinical data should all be developed as part of the postanalytic quality assurance process.

As per CLSI guidelines *(5)*, all patient laboratory records should be accessible and easily retrieved. They should be cross-referenced by both the patient name and by a second unique identifier (e.g., medical record member or laboratory accession number). The records are retained for a period of time as required by applicable federal or state regulations. CLIA'88 recommends that copies of patient records should be kept for a minimum of 10 yr, or as required by state law *(20)*. Electronic records are acceptable. However, guidelines for specimen retention time have not yet been agreed upon.

2.3.2. Timeliness of Reporting One of the important commitments in a laboratory's operation is providing test results within a time period suitable for prompt patient management decisions. This involves establishing limits for turnaround time (TAT) of tests offered. Turnaround time is defined as the time (e.g., number of hours or days) it takes to issue a result from receipt of the specimen in the laboratory to when the final report is resulted. What is acceptable will depend on the type of sample, the test requested, and the laboratory workload.

It is each laboratory's responsibility to check, at regular intervals, TAT against the threshold set for each of the assays offered. This can be monitored by establishing an indicator that would tabulate the percentage of reports completed by the stated TAT for the test(s). Prolonged TAT is to be investigated and corrective action taken perhaps by improving or changing the methodology. It is the laboratory director's responsibility to monitor TAT, including the volume of samples analyzed to detect trends, systematic errors, or local population variations that might have an effect on tests results or interpretation.

2.3.3. Correction of Errors Despite care and protective measures, errors do occur during the accessioning process. As stated in the CAP Molecular Pathology checklist, there must be a system for timely correction of both clerical and analytical errors. Clerical errors, such as incorrect spelling of a patient's name, incorrect address, wrong medical record number, or wrong date of birth, can occur during the accessioning of a patient specimen or as a result of misinformation provided by the referring clinician. These are to be corrected as soon as they become evident. Errors in the reported test results or misinterpretation of laboratory findings, must be corrected immediately. The referring physician is to be promptly notified and an amended report issued. A copy of this amended report is placed in the patient's file, including documentation of the error and the corrective action taken. To reduce clerical and analytical errors, laboratory results and patient demographics should be checked independently by a qualified member of the laboratory staff prior to signing out of the report by the laboratory director.

Errors occurring during the testing process, such as mislabeling a test tube, mixing two different patient specimens, or recording incorrect data on worksheets, must be corrected promptly. The technologist who has made the error(s) must document the event and indicate how it was corrected. In some cases, the test might have to be repeated at the laboratory's expense. In general, identifiable errors must be corrected immediately, because they could later translate into significant problems and potentially compromise the quality of patient care *(8)*.

2.3.4. Patient Confidentiality Records should be maintained in a manner to preserve patient confidentiality. All reports are issued to the referring physician or genetic professional, not the patient. They can be released to other healthcare professionals only with appropriate "authorization for release" from the patient. The laboratory must have a policy in place to protect the confidentiality of the test results reporting. For example, if a request for test results made by telephone from the referring physician's office, the laboratory personnel receiving the call must verify that it is the physician's office calling. The name and telephone number of the caller is obtained and patient's laboratory folder is reviewed to verify this is the referring physician. The event is documented and remains part of the patient file. If discrepancies exist, the matter should be referred to the laboratory director or manager.

Recently, new patient rights and healthcare provider regulations mandated by The Health Insurance Protability & Accountability Act (HIPAA) *(21)* were instituted. The regulations require all healthcare providers and their staff to restrict the use and disclosure of medical information and provide privacy rights for all their patients. Since each patient has the right to know who has access to and where their test results were sent, laboratories should incorporate into their QA policies how they are maintaining patient confidentiality. In New Jersey, the laboratory is legally required to share certain patient information, outside the realm of treatment, with agencies concerned with payment of services or healthcare, i.e., the New Jersey Department of Health and local health departments, that have not been authorized by the patient. To comply we developed a computer program to allow tracking of this information efficiently upon request of the patient.

3. PERSONNEL QUALIFICATIONS

Under the CLIA'88 rule, there is a change in the personnel standards for directors of high complexity laboratories. They must hold a doctoral degree in a chemical, physical, biological, or clinical laboratory science from an accredited institution and be certified and continue to be certified by a board approved by HHS *(2)*. The rule "grandfathers" in current experienced directors without board certification, and those who have served as directors of a laboratory performing high complexity testing prior to February 24, 2003 and have 2 yr of laboratory training and experience directing high complexity testing.

In addition, the American College of Medical Genetics requires the following to be met for laboratory directors *(3)*. They must have a PhD or MD, have at least 2 yr of experience or postdoctoral training in a clinical laboratory subspecialty, as well as being certified by specialty Boards. Board certification by the American Board of Medical Genetics or Canadian College of Medical Geneticists documents that this individual has both training and experience suitable for a director of molecular testing laboratory. The American Board of Pathology (ABP), the American Board of Clinincal Chemistry (ABCC), and the American Board of Bioanalysts (ABB) also provide recognized board certifications in this subspecialty. The laboratory director must be on site regularly or a accessible to the laboratory. He/she must ensure that the laboratory is in compliance with all regulations and maintains an ongoing quality assement program. The director determines the appropriate tests to be performed, the techniques to follow and the equipment and

Table 1
Indicator: Turn-Around Time

Important aspect of service to be evaluated	*Indicators to be monitored*	*Threshold for evaluation*
Turnaround time from specimen receipt to result reporting of B- and T-cell gene rearrangements done by the PCR method	The percentage of specimens for B- and T-cell gene rearrangements that are completed within 7 d of receipt	Less than 90%

Table 2
Plan to Monitor Turn-Around Time

Indicator to be monitored	*Source of collection*	*How collected*	*Who collects*	*Frequency of collection*	*Frequency of analysis*
The percent of specimens for B- and T-cell gene rearrangements done within 7 d of receipt.	From the data entered and stored in the laboratory computer system; the laboratory computer system indicates the date of receipt and the date of report for each patient sample.	The Molecular Statistics Test Log program is accessed from the computer system to provide a list of all accession numbers in which B- and T-cell gene rearrangement studies were resulted; the total number of cases of B-cell gene rearrangement assays and T-cell gene rearrangement assay are obtained.	The Molecular Statistics Log is printed by the Pathologist in Charge of the molecular laboratory (or designee).	On a monthly basis, the Section Chief of laboratory collects, tabulates, and reviews the data.	Statistic analysis is done biannually. (January to June and July to December) by the Section Chief and the Pathologist in Charge of molecular laboratory.

reagents to be used. He/she reviews, interprets, and reports all lab results and assures their accuracy.

ACMG also specifies qualifications for other laboratory personal such as the supervisor. They are to hold a Bachelors degree and have at least 3 years of experience in medical field. It is recommended that they obtain certification from such agencies as the American Society for Clinical Pathology (Molecular Pathology) or the National Credentialing Agency in which they could become Certified Laboratory specialist in Molecular Biology. Laboratory technologist must hold an undergraduate degree in a biological or chemical field or have at least five years of relevant laboratory experience.

The ability for laboratory personnel to troubleshoot assays and handle unexpected laboratory problems involves the continuous update of their knowledge. Continuing education is an important aspect of quality assement and improvement. This can be handled in several ways such as a journal club in which selected articles are discussed, reviewing results and interpretations of difficult and interesting patient reports, attending appropriate seminars and conferences. Such activities are to be documented for each laboratory member.

4. CONCLUSION

A rapidly increasing number of laboratories are now establishing molecular technologies to use for clinical diagnosis. Efforts by professional organizations such as ACMG, AACC, ABB, CAP, AMP, CLSI, and government agencies (exemplified in the new CLIA'88 rules) are being directed to the standardization of both the test systems and the laboratory procedure. Quality Assurance guidelines specific for molecular laboratories have not been clearly established yet. What has been presented in this chapter is intended to inform and guide such laboratory as to what to consider in instituting a QA program.

At our hospital, a plan has been in place for monitoring QA indicators for different laboratory sections. The purpose of the plan is to ensure that the department with all its available resources implements the mission of the hospital to provide the service needed to sustain the highest possible standard of patient care. The overall responsibility for the Quality Assessment plan lies with the Chairman of the department who shares reports of QA activity with the QA committees of the hospital. "Key indicators" for each laboratory section are selected at the beginning of each year and monitored monthly. If the indicator selected does not meet the established threshold, it is the responsibility of the director of that laboratory section to formulate an improvement plan and to implement it. The following is an example of one of the indicators (turnaround time) the Diagnostic Molecular Pathology laboratory is monitoring. It is divided into two parts: the first requires the selection of an indicator to be monitored (Table 1) and the second outlines a plan to monitor the indicator selected (Table 2).

Following are suggested key indicators that can be used to assess QA practices in your laboratory. See also the article by McGovern et al. *(6)* for other laboratory practice standards that can be used as indicators.

Suggested Indicators for the Preanalytical Phase

1. Percentage of specimens received without a requisition
2. Percentage of specimens received with inadequate demographics

3. Percentage of specimens with inadequate amounts of nucleic acid extracted
4. Percentage of specimens received with incorrect preservatives
5. Percentage of improperly stored specimens
6. Percentage of improperly labeled specimens
7. Percentage of rejected specimens (broken down by type)
8. Monitoring of the average time from specimen collection to receipt in laboratory
9. Percentage of clerical errors (broken down by type)

Suggested Indicators for the Analytical Phase

1. Average time from specimen receipt to verification of result
2. Percentage of tests that are repeated (broken down by reason)
3. Percentage of controls that are unacceptable
4. Percentage of contaminated runs

Suggested Indicators for Postanalytical Phase

1. Percentage of final reports that have significant clerical or typographic errors
2. Percentage of molecular diagnostic reports released after stated TAT
3. Percentage of molecular diagnostic reports confirmed as delivered electronically

Suggested Indicators for the Entire Testing Process

1. Percentage of correlations between two different methodologies
2. Percentage of a specific group of surgical pathology cases (hyperplastic lymph nodes, ASCUS, etc.) that were tested by a molecular method
3. Percent correlation of molecular results and surgical pathology findings
4. Percent correlation of retrospectively reviewed molecular diagnostic cases
5. Monitoring of the effect the test results has on patient care (i.e., chart review)
6. Average TAT for molecular diagnostic reports (from time of receipt in the laboratory to the time report is resulted)

REFERENCES

1. Medicare, Medicaid and CLIA Programs: Regulations Implementing the Clinical Laboratory Improvement Amendments of 1988 (CLIA). Final rule. *Fed. Regi.* **57:**7002, 1992. Available at http://www.fda.gov/cdrh/clia.
2. Medicare, Medicaid and CLIA Programs; Laboratory Requirements Relating to Quality Systems and Certain Personnel Qualifications; Final Rule (January 24, 2003) *Fed. Regi.* **68:**3640, 2003.
3. American College of Medical Genetics (ACMG). *Standards and Guidelines for Clinical Genetics Laboratories*. 3rd ed. Avalable at http://www.acmg.net.
4. College of American Pathology. Molecular pathology checklist. Avalable at http://www.cap.org.
5. NCCLS. *Molecular Diagnostic Methods for Genetic Diseases; Approved Guideline*. NCCLS document MM1-A. NCCLS, Wayne, PA, 2000.
6. McGovern, M., Benach, M., Wallenstein, S., Desnick, R., and Keenlyside, R. Quality assurance in molecular genetic testing laboratories. *JAMA* **281:**835–840, 1999.
7. Hofgartner, W. T. and Tait, J. F. Characteristics of clinical molecular-genetic testing laboratories in the United States. *Clin. Chem*. **45(8):**1288–1290, 1999.
8. Van Deerlin,V. Personal communication, 2003.
9. Beutler, E., Gelbart ,T., and Kuhl, W. Interference of heparin with the polymerase chain reaction. *BioTechniques.* **9:**166, 1990.
10. Ginocchio, C. C., Wang, X. P., Kaplan, M. H., et al. Effects of specimen collection, processing and storage conditions on stability of human immunodeficiency virus type 1 RNA levels in plasma. *J. Clin. Microbiol*. **11:**2886–2893, 1997.
11. Farkas, D. H., Drevon, A. M., Kiechle, F. L., DiCarlo, R. G., Heath, E. M., Crisan, D. Specimen stability for DNA-based diagnostic testing. *Diagn. Mol. Pathol*. **5(4):** 227–235, 1996.
12. Dicover, R. E., Herman, S. A., Saddiqu, K., Wafer, D., Dillon, M., and Bryson, Y. J. Optimization of specimen-handling procedures for accurate quantitation of levels of human immunodeficiency virus RNA in plasma by reverse transcriptase PCR. *J. Clin. Microbiol.* **36:** 1070–1073, 1998.
13. New York State Department of Health Laboratory, Wadsworth Center. Clinical Laboratory Evaluation Program, Laboratory Standards. Available at http://www.wadsworth.org/labcert/. (accessed February 2002).
14. NCCLS. *Clinical Laboratory Technical Procedure Manuals, 4th ed. Approved Guideline GP2-A4*.NCCLS Wayne, PA, 2002.
15. Gentra Systems. Purigene DNA Isolation Kit, Gentra Systems, Minneapolis, MN. Available at http://www.gentra.com.
16. Kwok, S. and Higuchi, R. Avoiding false positive with PCR. *Nature* **339:**237–238, 1989.
17. Longo, M. C., Berninger, M. S., and Hurtley, J. L. Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. *Gene* **93:**125–128, 1990.
18. Association for Molecular Pathology Statement. Recommendations for in-house development and operation of molecular diagnostic tests. *Am. J. Clin. Pathol*. **111:**449–463, 1999.
19. Shahangian, S., Holmes, E. H., and Taylor, R. Toward optimal PT use. *Med. Lab. Observ*. **32(4):**32–43, 2000.
20. Schwartz, M. K. Genetic testing and the Clinical Laboratory Improvement Amendment of 1988: present and future. *Clin. Chem*. **45(5):**739–745, 1999.
21. The Health Insurance Portability and Accountability Act of 1996 (Public Law 104-191). Standards for Privacy of Individually Identifiable Health Information–45 CFR Parts 160 and 164.

BIBLIOGRAPHY

1. Dequeker, E., Ramsden, S., Grody, W. W., Stenzel, T. T., and Barton, D. E. Quality control in molecular genetic testing. *Nat. Revi./Geneti.* **2:**717–722, 2001.
2. Diaz-Cano, S. PCR-Based alternative for diagnosis of immunoglobulin heavy chain gene rearrangement. *Diagn. Mol. Pathol*. **5(1):**3–9, 1996.
3. Holtzman, N. A. and Watson, M. S. *Promoting Safe and Effective Genetic Testing in the United States. Final Report of the Task Force on Genetic Testing.* National Academy Press, Washington, DC, 1997.
4. Mark, H. F. L. Surveying the environment for genetic testing: a regulatory perspective. *J. Assoc. Genet. Technol.* **27(4):**140–143, 2001.
5. McKenzie, A., and Balance, L. Everyday QA: a case study. *MLO* **35(3):** 24–29, 2003.
6. Neumaier, M., Braun, A. and Wagener, C., Fundaments of quality assessment of molecular amplification methods in clinical diagnostics. *Clin. Chem.* **44(1):**12–26, 1998.
7. Patton, S. and Stenhouse, S. Draft Best Practice Guidelines for Laboratory Internal Quality Control from EMQN Workshop on Internal Quality Control, Edinburgh, UK, 2001.

19 Verification of Molecular Assays

BRENT L. SEATON

1. INTRODUCTION

Before a molecular test is used to report patient results, it is the responsibility of each laboratory to determine (or verify) and document that the claims of the assay are accurate, reproducible, and sound. This verification process is established for each assay that a clinical laboratory performs, be it a molecular, serological, or microbiological assay. The verification process is not limited to in-house developed procedures ("home-brew" assays), but is required prior to the use of commercially available Food and Drug Administration (FDA)-approved assays or reagents as well. A well-designed verification will define the performance characteristics of the assay and the possible results generated by the assay.

2. VERIFICATION OF MOLECULAR TESTS

Standards have been determined by regulatory agencies for the verification and validation of molecular assays. In 1988, the US Congress passed the Clinical Laboratory Improvement Amendments (CLIA '88) to establish quality standards for all laboratory testing. The standards were designed to ensure the accuracy and reliability of human patient testing results regardless of where the test was performed. CLIA regulations are based on the complexity of the test performed. Laboratory testing is divided into simple or "waived" tests, tests of moderate complexity, and tests of high complexity. All molecular tests fall within the "high-complexity" category as defined by CLIA '88 *(1)*.

The College of American Pathologists (CAP) has established the Laboratory Accreditation Program to improve the quality of clinical laboratory services through voluntary compliance to established performance standards. CAP also has mandated that specific parameters must be verified for each assay before that assay is implemented into use in the laboratory *(2)*. Before CAP issues accreditation to a laboratory, the laboratory must show that it has defined the required verification parameters for the assays the lab performs.

As shown in Table 1, CLIA and CAP differ to some extent in their requirements for assay verification. CLIA differentiates assays into two groups: (1) FDA-approved tests and (2) non-FDA- cleared tests. The latter group would include commercial Research Use Only (RUO) tests, commercial kit components, and in-house developed assays. For FDA-cleared tests, CLIA requires that the performance specifications for accuracy, precision, and reportable range be determined or confirmed and that these performance specifications be comparable to those established by the manufacturer. The verification requirements for non-FDA-cleared tests are more extensive and include reference range, analytical sensitivity, and analytical specificity. The CAP requirements for verification do not differentiate among different types of assay, regardless of their FDA clearance status. CAP requires the more extensive verification process that CLIA requires for non-FDA-cleared tests.

Various professional organizations have published perspectives or guidelines concerning the verification of molecular assays. The National Committee for Clinical Laboratory Standards (NCCLS) has completed guidelines that address various types of molecular diagnostic assay *(3–9)*, as well as guidelines that address the evaluation of specific assay performance characteristics *(10–12)*. The American Society for Microbiology and the Association for Molecular Pathology have also published perspectives on the verification and use of molecular assays *(13,14)*.

3. REFERENCE RANGE

The simplest performance parameter to define is usually the reference range of the test. The reference range of a test is the range of values found in individuals who do not have the disease or condition that is being assayed by the test. Verification of the reference range is usually accomplished by assaying specimens from healthy donors to show that they produce a negative result. Given that it is much easier to find healthy individuals than those afflicted with a specific disease, the verification of this parameter is quite easy. Terms that might be used to define the reference range result are "not detected", "wild type," or "negative." For quantitative assays, the reference range will be reported as below a particular quantitative measurement (i.e., "less than 400 copies/mL"). For analytical tests (such as human leukocyte antigen [HLA] typing or HCV genotyping), each patient specimen being tested should possess the analyte being assayed. The testing laboratory could determine that the reference range is not applicable to the results of these types of assay.

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

Table 1
Assay Verification Requirements

	Regulatory agency or organization		
	CLIA		*CAP*
Assay type/clearance	FDA-cleared tests	All non-FDA cleared tests	All tests
Requirements for verification	Reportable range	Reportable range	Reportable range
	Precision	Precision	Precision
	Accuracy	Accuracy	Accuracy
		Reference range	Reference range
		Analytical sensitivity	Analytical sensitivity
		Analytical specificity[a]	Analytical interferences
		Other[b]	

[a]CLIA defines analytical specificity as "analytical specificity, including interfering substances."
[b]CLIA defines this parameter as "any other performance characteristic required for test performance."

When multiple specimen types are to be tested by the assay, the performance characteristics of the assay for each specimen type must be verified. An example is an assay that is designed to detect the presence of cytomegalovirus (CMV). Cytomegalovirus can be detected in whole blood, cerebrospinal fluid (CSF), and amniotic fluid by molecular methods *(15)*. If the assay is to be used by the laboratory to detect CMV in any of these three specimen types, the reference range must be determined for each specimen type independently. This requirement for verification of all specimen types to be used in an assay is applicable for all other performance parameters to be verified.

4. ANALYTICAL SENSITIVITY

Analytical sensitivity refers to the lowest measurable amount of a target nucleic acid that can be detected (qualitative tests), quantified (quantitative tests), or characterized (analytical tests) reproducibly by the assay. The term "detection limit" is often used interchangeably with "analytical sensitivity." The measurement used to define the analytical sensitivity of a test will depend on the type of assay. For assays designed to detect or quantify a particular infectious pathogen in a specimen, the units of measure might be "copies/mL," "colony-forming units/mL," or "plaque-forming units/mL." For assays designed to detect a specific genetic mutation, the analytical sensitivity might be expressed as a percentage. The analytical sensitivity can be reported as 0.1% if the assay can detect one mutant allele in the presence of 999 wild-type alleles.

In each case, analytical sensitivity can be determined by testing specimens with known quantities of the nucleic acid analyte being assayed. Clinical specimens containing quantified target analyte (or a "normal" specimen spiked with a known quantity of the analyte) are serially diluted with normal specimens, or specimen matrix, and assayed to determine the detection limit of the assay. Multiples of each dilution series should be assayed to define the amount or percentage of target analyte that can be detected with 100% confidence. Laboratorians should remember that the analytical sensitivity of an assay refers to the sensitivity of the whole assay, not just the amplification and detection part of the assay. The complete assay procedure, including specimen preparation, must be performed to accurately define analytical sensitivity. For assays in which multiple specimen types can be assayed, the analytical sensitivity must be defined for each specimen matrix to be used in that assay.

5. ANALYTICAL SPECIFICITY

The analytical specificity of an assay is defined as the degree by which the assay detects the specific target yet does not produce a positive result in the presence of nonspecific targets. The analytical specificity of the assay is assessed to verify that the assay only detects the target it claims to detect or analyze and does not crossreact with related targets. In infectious disease testing, one must show that the assay is able to discriminate between the target pathogen and related pathogens. Also, the assay must be able to differentiate between the target pathogen and other pathogenic or commensal flora that might be found in the specific specimen type(s) to be tested by the assay. In genetic testing, it must be documented that the assay can discriminate between the intended genetic sequence and related genes, pseudogenes, or alleles.

If multiple specimen types are to be assayed using the molecular test, specificity of the assay should be verified using all organisms that can potentially be found in the various specimen types. Table 2 gives an example of the various pathogens one would want to test when assessing the analytical specificity for an assay that is designed to detect CMV in either CSF or amniotic fluid. The list of potential pathogens causing meningitis differs from the list of potential pathogens in amniotic fluid, and both groups of pathogens must be tested to verify the specificity of the assay.

Included in the requirement for verification of analytical specificity by CLIA is the requirement for verification of potential interfering substances. A number of studies have identified potential inhibitors to the polymerase chain reaction (PCR), and specimens containing these interfering substances should be avoided *(16–21)*. Most molecular assays will incorporate an inhibitory control that verifies the absence of amplification inhibitors in the extracted nucleic acids of the samples tested. The inhibitory control (sometimes referred to as the "internal control") can be a nucleic acid target inherent to the specimen or can be exogenously added to the specimen prior to nucleic acid extraction. The inhibitory control is extracted and amplified in parallel with the analyte nucleic acid target. Detection of the amplified internal control signifies successful nucleic acid extraction and successful amplification. Documentation of the usage of this control during the verification procedure should fulfill the requirement for verification of this assay parameter.

Table 2
Analytical Specificity of a CMV Assay

Specimen type	*Cerebrospinal fluid*	*Amniotic fluid*
Clinical syndrome associated with infection of specimen type	Meningitis	Prenatal infection
Related herpesviruses	HSV-1, HSV-2, VZV, EBV, HHV-6, HHV-7, HHV-8	HSV-1, HSV-2, VZV, EBV, HHV-6, HHV-7, HHV-8
Other pathogens possibly causing clinical syndrome	Arboviruses Enteroviruses *Haemophilus influenzae* *Streptococcus pneumoniae* *Neisseria meningitidis* *Mycobacterium tuberculosis* *Ricketssia* spp.	*Toxoplasma gondii* Parvovirus B19 Rubella virus Papillomaviruses
Other pathogens possibly found in specimen type	JC virus HIV-1 *Toxoplasma gondii*	

Note: This Table 2 demonstrates the different requirements when verifying analytical specificity of an assay that can be used for multiple specimen types. CMV infection of different body fluids (CSF vs amniotic fluid) or organs can lead to different clinical manifestations. Therefore, different groups of pathogens will need to be analyzed to verify assay specificity when multiple specimen types are accepted for testing of a particular assay.

6. REPORTABLE RANGE

The reportable range of an assay is the range of values that will generate a positive result for the specimens assayed by the test procedure. Clinical patient specimens, positive for the given nucleic acid analyte, are assayed to determine the values expected to be produced by clinically positive specimens. Depending on the type of assay (qualitative, quantitative, analytical), the reportable range can have multiple definitions.

For qualitative tests that are only detecting the presence of a molecular analyte, the reportable range would be that range of values that give a positive result. The reference range and analytical sensitivity have already been quantified to determine the cutoff values used to determine positive and negative results. Clinical specimens from patients that are positive for the disease condition are tested by the assay to verify that the values generated by such specimens are above the cutoff for the assay.

For quantitative assays, the reportable range will represent the linear range of quantitation for the assay. The linear range is the span of analyte concentration for which the final output value of the system is directly proportional to the analyte concentration. As with the case of analytical sensitivity, linear range must be kept in context of the whole assay, not just the amplification and detection steps. The complete assay procedure, including specimen preparation, must be performed to accurately define the linear range of results. The linearity of an assay is determined by testing different concentrations of an analyte in the appropriate specimen matrix. Known amounts of the molecular analyte can be spiked into the appropriate specimen matrix. Conversely, a specimen with an elevated amount of the analyte can be serially diluted with analyte-negative specimen matrix. After assaying the specimens, a plot of the logarithmically transformed data is performed in which the system output values are plotted against starting analyte concentrations. The range in which the plotted data conforms to a straight line defines the linear range of the assay. Guidelines for the design and data evaluation of experiments to determine linear range can be found in the current edition of NCCLS documents EP-6 (*Evaluation of the Linearity of Quantitative Analytical Methods) (11)* and MM-6 *(Quantitative Molecular Methods Infectious Diseases) (7).*

For analytical assays, the reportable range will define the alleles, mutations, or genotypes that can be detected and differentiated by the assay. The ability of the assay to detect and differentiate among these analytes must be verified and documented.

7. PRECISION

Precision refers to the ability of an assay to produce the same result for a given sample when repeatedly tested over time (reproducibility). There are three levels of precision as it pertains to routine laboratory testing: (1) *within-run precision* is the measurement of reproducibility when multiple aliquots of a sample are analyzed in parallel during one assay run; (2) *between-run precision* is the measurement of reproducibility when multiple aliquots of a sample are analyzed in separate assay runs on the same day; and (3) *between-day precision* is the measurement of reproducibility when multiple aliquots of a sample are analyzed over a period of more than 1 day. Depending on the anticipated testing volume for the assay, a laboratory might or might not be required to define each type of precision.

Given that a large number of sample replicates will be needed to accurately define assay precision, the laboratory should carefully plan the experiments needed to define this parameter. A large volume of the test material should be prepared and then divided into multiple aliquots, which will then be assayed independently. Precision testing should also include more than one lot or batch of test reagents. For qualitative tests, at least two levels of the analyte should be tested (negative and low positive). For quantitative tests, specimens with analytes in the low, moderate, and high ranges of quantitation should be selected for precision verification as well as a negative sample. For assays that are analytical in nature (tests looking at genetic variants or mutations), the number and types of specimens to use for precision studies are less well defined. At the least, samples

Table 3
Commercial Sources of External Standards/Control Materials

Company	*Website and telephone number*
AcroMetrix	www.acrometrix.com; 888-746-7921
Advanced Biotechnologies	www.abionline.com; 800-426-0764
American Type Culture Collection	www.atcc.org; 800-638-6597
BBI Diagnostics	www.bbii.com; 800-676-1881
National Institute of General Medical Sciences Human Genetic Cell Repository	locus.umdnj.edu/nigms; 800-752-3805

containing the mutations/variants that are found in the majority of the target population should be used to define precision. Guidelines for experiment design, data collection, and data evaluation can be found in the current edition of NCCLS document EP5 (*Evaluation of Precision Performance of Clinical Chemistry Devices) (10).*

8. ACCURACY

Accuracy is one of the last performance parameters to be defined for the assay. By this time, the values for positive and negative results have been defined and the assay has been shown to be specific and reproducible for detecting the analyte(s) for which it was developed. It is now time to take the assay out for a test run.

Accuracy refers to the ability of the test to produce a correct result compared with an external standard. External standards can be previously tested patient samples from interlaboratory exchanges or proficiency surveys. The standards can also be target nucleic acids or viruses spiked into the appropriate specimen matrix. Table 3 lists commercial sources for standards that might be used for accuracy studies.

To verify the accuracy of an assay, a panel of specimens is organized and assayed in a blind manner. For qualitative assays, a panel of known positive and negative samples from each specimen type should be assayed. If possible, positive specimens with low, moderate, and high amounts of the analyte should be included in the panel. For quantitative tests, a panel of samples of known analyte quantity (including samples with no analyte) should be used to verify accuracy. The analyte quantity in the specimens to be tested should span the linear range of the assay. For analytical tests, a panel of samples with known alleles, mutations, or genotypes should be used to verify their respective analytical tests.

9. OTHER PARAMETERS

CLIA ’88 states that “any other performance characteristic required for test performance” must also be verified or confirmed. Given the generality of this statement, this requirement can be interpreted in a myriad of different ways. Specimen or sample stability is one assay parameter that the laboratory might want to verify. The need for verification of this parameter depends on the location of the testing laboratory with respect to the patients from which the specimens are obtained. For laboratories in the same physical location as the patients, specimen stability might not be a major concern. However, for central or referral laboratories, optimal specimen storage and transport conditions should be identified to ensure the validity of the assay results with respect to the patient. Simple experiments can be designed to test the length of time that specimens can be stored at various temperatures and still allow for accurate detection of the nucleic acid analyte in these specimens.

10. CLINICAL VALIDATION

Clinical validations are intended to show that the results produced by an assay correlate with clinical disease. Neither CLIA ’88 nor CAP requires the clinical validation of an assay prior to implementation of its use in a testing laboratory. Although clinical validations are not required before testing, these types of validation study are still needed before the assay can be used for clinical diagnosis. In many instances, clinical studies for a given analyte have been performed and published and the results of these studies can be extended to similar assays developed or introduced elsewhere. However, care must be taken in interpreting these findings. Because of lack of standardization among most molecular assays (especially home-brew assays), the results from one assay are not necessarily equivalent to a similar assay performed in a different laboratory.

REFERENCES

1. Medicare, Medicaid and CLIA programs: regulations implementing the Clinical Laboratory Improvement Amendments of 1988 (CLIA), *Fed. Regi*. 1992.
2. College of American Pathologists Laboratory Accreditation Program. Internet website http://www.cap.org.
3. National Committee for Clinical Laboratory Standards. *Molecular Diagnostic Methods for Genetic Diseases*; *Approved Guideline*. NCCLS document MM1-A. NCCLS, Wayne, PA, 2000.
4. National Committee for Clinical Laboratory Standards. *Immunoglobulin and T-cell Receptor Gene Rearrangement Assays*; *Approved Guideline*. NCCLS document MM2-A2. NCCLS, Wayne, PA, 2002.
5. National Committee for Clinical Laboratory Standards. *Molecular Diagnostic Methods for Infectious Diseases*; *Approved Guideline*. NCCLS document MM3-A. NCCLS, Wayne, PA, 1995.
6. National Committee for Clinical Laboratory Standards. *Nucleic Acid Amplification Assays for Molecular Hematopathology*; *Approved Guideline*. NCCLS document MM5-A. NCCLS, Wayne, PA, 2000.
7. National Committee for Clinical Laboratory Standards. *Quantitative Molecular Methods for Infectious Diseases; Approved Guideline*. NCCLS document MM6-A. NCCLS, Wayne, PA, 2001.
8. National Committee for Clinical Laboratory Standards. *Nucleic Acid Sequencing Methods in Diagnostic Laboratory Medicine; Approved Guideline.* NCCLS document MM9-A. NCCLS, Wayne, PA, 2004.
9. National Committee for Clinical Laboratory Standards. *Genotyping for Infectious Disease: Identification and Characterization; Proposed Guideline.* NCCLS document MM10-P. NCCLS, Wayne, PA, 2005.
10. National Committee for Clinical Laboratory Standards. *Evaluation of Precision Performance of Clinical Chemistry Devices; Approved Guideline.* NCCLS document EP5-A. NCCLS, Wayne, PA, 1999.
11. National Committee for Clinical Laboratory Standards. *Evaluation of the Linearity of Quantitative Analytical Methods; Approved Guideline.* NCCLS document EP6-A. NCCLS, Wayne, PA, 2001.
12. National Committee for Clinical Laboratory Standards. *Interference Testing in Clinical Chemistry; Approved Guideline.* NCCLS document EP7-A. NCCLS, Wayne, PA, 2003.
13. McCurdy, B.W. (ed.). *Cumitech 31: Verification and Validation of Procedures in the Clinical Microbiology Laboratory*. ASM, Herndon, VA, 1997.

14. Association for Molecular Pathology Statement.Recommendations for in-house development and operation of molecular diagnostic tests. *Am. J. Clin. Pathol.* **111:**449–463, 1999.
15. Mansy, F., Brancart, F., Liesnard, C., Bollen, A., and Godfroid, E. A PCR based DNA hybridisation capture system for the detection of human cytomegalovirus. A comparative study with other identification methods. *J. Virol. Methods* **80:**113–122, 1999.
16. Akane, A., Matsubara, K., Nakamura, H., Takahashi, S., and Kimura, K. Identification of the heme compound copurified with deoxyribonucleic acid (DNA) from bloodstains, a major inhibitor of polymerase chain reaction (PCR) amplification. *Forensic Sci*. **39:**362–372, 1994.
17. Satsangi, J., Jewell, D.P., Welsh, K., Bunce, M., and Bell, J.I. Effect of heparin on polymerase chain reaction. *Lancet* **343:**1509–1510, 1994.
18. Wang, J.T., Wan, T.H., Sheu, J.C., Lin, S.M., Lin, J.T., and Chen, D.S. Effects of anticoagulants and storage of blood samples on efficacy of the polymerase chain reaction assay for hepatitis C virus. *J. Clin. Microbiol*. **30:**750–753, 1992.
19. Wiedbrauk, D.L., Werner, J.C., and Drevon, A.M. Inhibition of PCR by aqueous and vitreous fluids. *J. Clin. Microbiol*. **33:**2643–2646, 1995.
20. Rossen, L., Noskov, P., Holmstrom, K., and Rasmussen, O.F. Inhibition of PCR by components of food samples, microbial diagnostic assays and DNA-extraction solution. *Int. J. Food Microbiol*. **17:**37–45, 1992.
21. Khan, G., Kangro, H.O., Coates, P.J., and Heath, R.B. Inhibitory effects of urine on the polymerase chain reaction for cytomegalovirus DNA. *J. Clin. Pathol*. **44:**360–365, 1991.

20 Standards and Standardization of Molecular Diagnostics

JOHN P. JAKUPCIAK AND CATHERINE D. O'CONNELL

1. INTRODUCTION

DNA has become a major target for clinical laboratory testing over the past 5 yr, and RNA testing is emerging for infectious disease and gene expression *(1)*. To normalize laboratory results across different technology platforms as well as between laboratories, standardized reagents will become increasingly important. Reliable standards promote the speed at which a diagnostic test can be offered, as well as third-party reimbursement. Standardized control reagents ensure the ability of diagnostic laboratories to pass proficiency testing and quality assurance/quality control (QA/QC) measurements *(2)*. Consensus guidelines endorsed by professional societies and governmental agencies provide a framework for determining standardization needs.

In its role supporting US science and industry, the National Institute of Standards and Technology (NIST), a nonregulatory agency of the US Department of Commerce, provides physical and chemical standards in support of national commerce, manufacturing, and science *(3)*. These materials are available internationally as Standard Reference Materials (SRMs) for use by industry developing assays and/or technology platforms for diagnostic use, by regulatory agencies ensuring the quality and efficacy of these assays, and by clinical laboratories providing diagnostic tests for patients. Traditionally, NIST responds to standard needs as defined by these communities. Consensus is developed through NIST workshops attended by representatives of these communities as well as direct request by other governmental agencies. Specific examples of ongoing programs within the Biotechnology Division at NIST are described in this chapter.

2. TRINUCLEOTIDE REPEAT MEASUREMENTS

Basic research and technological advances in human genetics, biochemistry, and model systems have brought much progress toward the understanding of human infectious, hereditary, and somatically acquired diseases. In fact, whereas in 1991, triplet repeat expansion diseases could be described in a single article, now the remarkable developments within each disease has created volumes of work *(4,5)*. One factor contributing to triplet repeat diseases is the adoption of unusual non-B DNA structures by the repetitive elements. These structures are associated with disease progression and severity. Despite all of the advances, an effective therapy is yet unrealized. However, the defining element of these diseases, the unstable DNA mutation, can accurately be measured. Some diseases already have diagnostics methods available [i.e., Huntington's disease and Fragile X syndrome *(6,7)*]. Yet, in the case of Fragile X syndrome, accurate size determination of the triplet repeat region is not easily determined because of the nature of the repeat elements associated with this disease; consisting entirely of the C and G nucleotides. This sequence content increases the error rate of polymerase chain reaction (PCR) amplification and is commonly methylated in the disease state.

2.1. MEASUREMENT TECHNOLOGIES FOR ACCURATE SIZING: FRAGILE X SYNDROME Current measurement technologies for repetitive elements and their adjacent flanking sequences are Southern blot hybridization, PCR amplification and electrophoretic separation, and DNA sequencing *(7)*. The need for reference materials for Fragile X syndrome was discussed at a workshop held at NIST in 1998 entitled "Standards for Nucleic Acid Diagnostic Applications" *(8)*. Reference materials for this syndrome are needed to provide accurate repeat size measurements across technology platforms and interlaboratory diagnostic and prognostic agreement. The specific size range for such a standard reference material was based on the likelihood of full mutation transmission *(9)*.

The Biotechnology Division established a measurement program in this area, focusing on accurate size measurements after PCR amplification and sequencing. The accuracy of an optimized PCR amplification protocol to correctly measure the number of $(CGG)_n$ repeats from normal, gray zone, and premutation length alleles was determined *(4)*. The DNAs used in this study were reported to contain CGG repeat elements ranging from 29 to 110 repeats. Both slab-PAGE (polyacrylamide gel electrophoresis) and capillary measurements were conducted, and the factors impacting sensitivity, accuracy, and reproducibility of

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

Table 1
Measurement of (CGG) Repeats

Coriell No.	*Allele size(s): slab-PAGE*	*Peak ratio*	*Allele size(s): CE*	*Peak ratio*
6906	94	—	90	—
6910	29,88	0.082	29,85	0.06
6968	32,109	0.59	32,105	0.25
7541	29,31	1.05	28,30	1.03
13664	29,51	0.80	29,51	0.79

results were examined. Long repeats (full mutation length alleles), which typically have clear clinical phenotypes, were not included in this study. DNA obtained from both previously tested clinical specimens as well as cell lines obtained from the Coriell Cell Repository were analyzed after PCR amplification. These samples were used as in-house controls by two clinical laboratories. Our initial measurements focused on measurement variability: (1) between slab-PAGE and capillary (CE) separation systems, (2) interlane variability, (3) intergel variability (slab-PAGE), and (4) variability during PCR amplification. We performed statistical analyses on system reproducibility and interlane and intergel variability. Samples were run in triplicate for all measurements and the analysis was performed using GeneScan(™) analysis software. DNA sequencing was performed to verify repeat lengths.

2.1.1. Slab-PAGE Analysis As expected, the shorter alleles were more easily amplified and sequenced than longer alleles. The standard deviations for interlane measurements in slab-gels ranged from 0.05 to 0.35. The variation in size measurements performed on different gels and PCR amplifications ranged from 0.06 to 0.30 *(7)*. This suggests that these measurements varied by up to a single nucleotide (0.33 of a three-nucleotide repeat).

2.1.2. CE Analysis The CGG repeat measurements performed by capillary electrophoresis were slightly more precise, with standard deviations ranging from 0.02 to 0.29. However, allele sizes observed after CE separations were significantly smaller than those obtained after slab-gel electrophoresis (Table 1, samples 6910 and 6968). DNA sequence analysis confirmed that the size measurements were correct for the slab-gel data and inaccurate for the CE results. It was hypothesized that the proprietary gel matrix used for capillary electrophoresis (POP-4(™)) leads to anomalous rapid electrophoretic mobility of CG-rich sequences *(10)*.

2.1.3. Allele Size Analysis The peak ratio of each allele was compared within each female sample (Table 1). The detection method with peak ratios closest to 1.00 is more accurate but not necessarily more precise. The results indicate a gradual decrease in peak ratio with increase of allele size. As shown, premutation alleles contain the highest peak ratio discrepancy (comparison of samples 6910 and 6968 with 7541 and 13664). This bias would be even more pronounced in the detection of larger, premutation and full mutation length alleles *(7)*.

Table 1 also shows that CE analysis has greater error detecting the presence of long alleles than slab-PAGE, as shown by the difference between peak ratios (i.e., shaded areas samples 6968; 0.59 for slab-PAGE vs 0.25 for CE). This suggests that the electrokinetic injection used in CE results in a bias toward capillary loading of smaller alleles obtained from female specimens. Therefore, both amplification and amplicon loading appear to contribute to peak ratio discrepancy.

2.2. SUMMARY OF ACCURATE SIZING METHODS FOR FRAGILE X SYNDROME The accuracy of our sizing data for Fragile X measurements within the normal, gray zone, and premutation allele sizes was within one repeat length for slab-PAGE measurements. The precision was equally high in lane-to-lane comparisons, comparisons of PCR results between gels, and in multiple PCR amplifications. There is no statistically significant evidence for heterogeneity of size determination after separation by either slab-PAGE or CE measurements. Hence, the gel matrix and running conditions for slab-PAGE were suitable for accurate size determinations (as confirmed by DNA sequencing) despite known migration anomalies *(10)*. A GC-rich sizing standard would improve the accuracy of sizing by capillary electrophoresis, as this separation method results in high-precision measurements but incorrect size determinations.

Our data reveal several important considerations in the performance of Fragile X testing by PCR. First, size measurements were not directly comparable between the two separation systems for the larger, premutation, and presumably full mutation length alleles. Second, the POP4 polymer used in this study resulted in premutation size measurements that varied from actual size (as measured by DNA sequencing) by four to eight repeat elements. The greatest error (three repeats) detected in slab-PAGE measurements was found in the sample with the longest allele size—112 repeats as determined by DNA sequencing. All other normal, gray zone, and premutation measurements agreed with DNA sequencing. In addition, electrokinetic injection by CE resulted in allele-biased loading of premutation alleles. Because PCR methods are developed that robustly amplify full mutation alleles, the bias for loading smaller alleles could impact the ability to detect these alleles. High-precision measurements for samples containing long premutation and full mutation length alleles are currently under validation. Sizing standards that effectively allow cross-platform and interlaboratory comparisons are under development as a NIST SRM.

3. MEASUREMENT TECHNOLOGIES FOR MUTATION DETECTION: TP53 STANDARDS

According to the latest information provided by the publicly funded GeneTests website (http://www.genetests.org/), 928 diseases are currently tested for in medical genetics laboratories. Most of the current methods for the detection of mutations use PCR amplification technologies with subsequent analyte detection after electrophoretic separation or chromatographic separation of the products *(11–15)*. Microarray technologies have also been used to detect DNA mutations *(14)*. The US Food and Drug Administration (FDA) evaluates many of the innovations for commercial development of diagnostic tests. Although some have received approval, molecular diagnostic kits are rare. FDA approval of these technologies has been hampered, in part, by the lack of universal standards. New diagnostic assays emerge each year and their validation as accurate measurement technologies is critical for establishing the assay

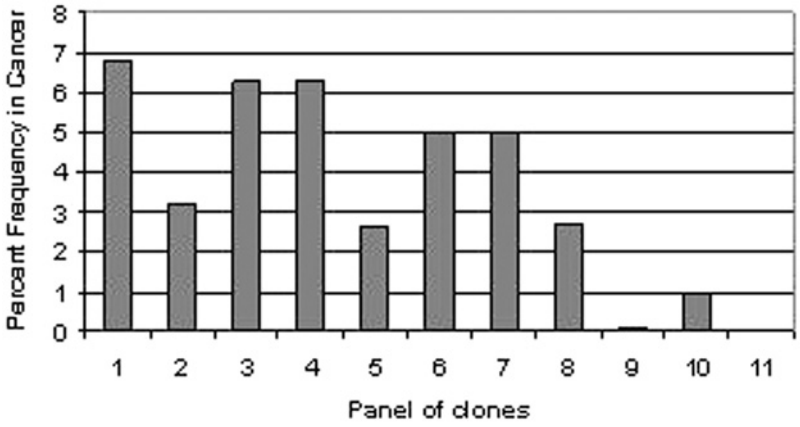


Fig. 1. Most common TP53 point mutations in cancer. Shown is the frequency (in percent) of the mutation contained in each clone as reported to occur in various cancers. For example, the mutation of clone 1 occurs in 6.8% of all cancers associated with mutations in the TP53 gene (or the mutation occurs in 1200 of the almost 18,000 reported cases). Clones 9–11 represent less than 1% of mutations associated with cancer.

Table 2
Summary of Location and Nature of Mutations in the Panel

Clone	*Amino acid position*	*Type of mutation*
1	175	G to A
2	245	G to A
3	248	C to T
4	248	G to A
5	249	G to T
6	273	C to T
7	273	G to A
8	282	C to T
9	128	C to T
10	237	T to C
11	328	T to C

as a routine test. Standards provide a reliable way to both monitor the sensitivity of mutation detection and to discern possible introduction of errors that potentially occur during the amplification and separation procedures *(16–18).* Mutations in the TP53 gene are the single most common genetic alterations observed in human cancers *(19,20).* For this reason, mutation detection of TP53 was selected as a candidate by NIST to create standards. These are expected to assist health care industries in the validation of new cancer detection assays independent of technology platform. The detection of mutations within this gene, like many others, is a difficult problem because mutations occur throughout the gene *(19).* In addition, the status of the TP53 gene in cancer has been linked to poor clinical outcome. TP53 mutations have been demonstrated as predictive indicators of recurrence and death in breast cancer and response to chemotherapy *(21).* Hence, TP53 genetic assays have the potential to become clinical diagnostic tools to track tumor progression and to determine therapy *(22,23).* Testing for TP53 mutations in Li–Fraumeni syndrome, although not common, is currently conducted in the United States by seven clinical laboratories (Genetests Laboratory Directory).

A panel of 12 plasmid clones has been developed by NIST that contains a 2.0-kb region of the TP53 gene spanning exons 5–9 (Fig. 1). These materials, as well as all NIST SRMs, will provide the clinical community with traceability between laboratory controls and a common reference material, thus providing interlaboratory conformity. Eleven of these clones contain a single mutation within the mutational hot spots of the TP53 gene (Table 2). The twelfth is wild type for this region of the gene. The eight most common single-base substitution mutations in human cancer are represented in this panel (clones 1–8), as well as three that proved difficult to detect (clones 9–11) by various scanning technologies *(24–25).* Each clone has been fully sequenced on both strands of the TP53 region.

To determine the effectiveness of this panel to serve as a SRM, we analyzed the single-point mutations by capillary electrophoresis–single-strand conformational polymorphism (CE-SSCP), denaturing gradient gel electrophoresis (DGGE), and denaturing high-performance liquid chromatography (DHPLC). The detection capabilities of these technologies were compared.

To this end, we have determined the accuracy of each method in detecting the mutations in our TP53 SRM panel in comparison to direct sequence analysis. CE-SSCP methods could detect 9 of the 11 mutations in the panel, a detection sensitivity of 82%. This is within the range of sensitivity reported by others *(26).* In this regard, CE-SSCP cannot be used to rule out the presence of a mutation, but remains a simple, inexpensive screening tool. DGGE is expected to detect mutations with a sensitivity of about 95% *(27)* and it successfully detected all of the mutations in the SRM panel. Although DGGE detected all of the mutations, unlike SSCP-CE it is not readily adaptable for high throughput. Similarly, DHPLC has been described as having a sensitivity of 95–100% *(28)* and it detected all of the mutations in the SRM panel. Once the appropriate assay conditions and column design are in place, high-throughput analysis is possible using DHPLC because of its quick run time. On the horizon is the use of microarrays for mutation detection *(14,29).* Their strength is in the detection of single-nucleotide substitutions (88% sensitivity) but are less useful for insertions and deletions *(29).*

Each of the three technologies has limitations in sensitivity to detect mutations that are not included in the SRM panel. This is an issue when screening disease genes that have many mutations (i.e., the TP53 gene). The SRM TP53 panel can accurately determine a method's sensitivity of detection of specific mutations

and would be useful for the development of new mutation detection technologies. As a large area of the TP53 gene was used in the development of this panel of clones, position and sequence context effects can be evaluated through the use of different primer sets.

4. SUMMARY OF MUTATION DETECTION IN TP53

The SRM panel of TP53 mutations fulfills the NIST mission statement of providing reliable, traceable, accurate measurements or standards. In addition, this SRM can normalize data across different technological platforms, so that cancer mutation scanning information can be readily shared and easily compared among different diagnostic laboratories using the same or different analytical techniques. This panel can be used to validate-in house laboratory controls. These in-house controls could change with respect to production date (cell lines) or depletion of previously tested samples in laboratories using these materials for controls. The use of common reference materials will lower the inherent error of the measurement techniques and interlaboratory variation by ensuring that protocol optimization can always detect the SRMs. Hence, each diagnostic lab can adjust its protocol(s) to obtain an expected and reproducible results independent of the method employed *(1)*. Equally important, the FDA can use data normalized against SRMs in its regulatory response and potentially facilitate the approval of drug therapies and/or new diagnostic technologies.

Our work compared three different established methods for the detection of single-point mutations *(16)*. It will be useful to help make correct choices of methods and conditions for the analysis of a great number of malignant, neurological, immunological, and other diseases or for direct studies of genomic variation. The value of this information is not limited to detection of mutations of TP53, but it has implications for the analysis of other genes as well.

REFERENCES

1. Barker, P. E., Watson, M. S., Ticehurst, J. R., Colbert, J. C., and O'Connell, C. D. NIST physical standards for DNA-based medical testing. *J. Clin. Lab. Anal.* **16:**5–10, 2002.
2. American College of Medical Genetics. *Standards and Guidelines for Clinical Genetics Laboratories,* 2nd ed., American College of Medical Genetics, Bethesda, MD.
3. Schooley, J. F. *Responding to National Needs: The National Bureau of Standards Becomes the National Institute of Standards and Technology.* 1969–1993.: US Government Printing Office Washington, DC, 2000.
4. Wells, R. D. and Warren, S. T. (eds.). *Genetic Instabilities and Hereditary Neurological Disease.* Academic, San Diego, CA, 1998.
5. Jakupciak, J. P. and Wells, R. D. Genetic instabilities of triplet repeat sequences by recombination. *Life* **50:**355–359, 2000.
6. Evers-Kiebooms, G., Harper, P., Zoeteweij, M., et al. Predictive DNA-testing for Huntington's disease and reproductive decision making: a European collaborative study. *Eur. J. Hum. Genet.* **10:** 167–176, 2002.
7. O'Connell, C. D., Richie, K. L., Jakupciak, J. P., Amos, J., and Atha, D.H. Standardization of PCR amplification for fragile X trinucleotide repeat measurements. *Clin. Genet.* **61:**13–20, 2002.
8. O'Connell, C. D. Standards for Nucleic Acid Diagnostic Applications, NIST, Gaithersburg, MD, 1998.
9. Emery, A. E. (ed.) *Principles and Practice of Medical Genetics.* Churchill Livingstone, Edinburgh, UK, 1997.
10. Chastain, P. D., Eichler, E. E., Kang, S., Nelson, D. L., Levene, S. D., and Sinden, R. R. Anomalous rapid electrophoretic mobility of DNA containing triplet repeats associated with human disease genes. *Biochemistry.* **34:**16,125–16,131, 1995.
11. Eng, C., Brody, L. C., Wagner, T. M. U., et al. Interpreting epidemiological research: blinded comparison of methods used to estimate the prevalence of inherited mutations in BRCA1. *J. Med. Genet.* **38:**824–833, 2001.
12. Kristensen, V. N., Kelefiotis, D., Kristensen, T., and Borresen-Dale, A-L. High-throughput methods for detection of genetic variation. *Biotechniques* **30:**318–332, 2001.
13. Ellis, L. A., Taylor, C. F., and Taylor, G. R. A comparison of fluorescent SSCP and denaturing HPLC for high throughput mutation scanning. *Hum. Mutat.* **15:**556–564, 2000.
14. Cotton, R. G. H. Mutation detection 2001: novel technologies, developments and applications for analysis of the human genome. *Hum. Mutat.* **19:**313–314, 2002.
15. Cotton R. G. H. Slowly but surely towards better scanning for mutations. *Trends Genet.* **13:**43–45, 1997.
16. O'Connell, C. D., Tully, L., Devaney, J. M., Marino, M., Jakupciak, J. P., and Atha, D. H. Renewable standard reference material for the detection of TP53 mutations. *Mol. Diagn.* **7:**85–97, 2003.
17. O'Connell, C. D., Barker, P. E., Marino, M., et al. Molecular biomarkers used to detect cellular/genetic damage in tissue engineered skin, Schuttle, E. and Picciolo, G. L., eds., *Tissue Engineered Medical Products.* ASTM, West Conshohocken, PA, 2003.
18. Sunar-Reeder, B., Aydemir, S., Khan, A. R., et al. Use of TP53 reference materials to validate mutations in clinical tissue specimens by single-strand conformational polymorphism analysis. *Mol. Diagn.* **8:**123–130, 2004.
19. Hollstein, M., Sidransky, D., Vogelstein, B. and Harris, C.C. P53 mutations in human cancer. *Science* **253:**49–53, 1991.
20. Vogelstein, B. and Kinzer, K. p53 function and dysfunction. *Cell* **70**:523–526, 1992.
21. Somlo, G., Simpson, J. F., Frankel, P., et al. Predictors of long-term outcome following high-dose chemotherapy in high-risk primary breast cancer. *Br. J. Cancer* **87:**281–288, 2002.
22. Liu, M. C. and Gelmann, E. P. p53 gene mutations: case study of a clinical marker for solid tumors. *Semin. Oncol.* **29:**246–257, 2002.
23. Geisler, S., Lonning P. E., Aas, T.,et al. Influence of TP53 gene alterations and c-erb B-2 expression on the response to treatment with doxorubicin in locally advanced breast cancer. *Cancer Res.* **61:**2505–2512, 2001.
24. Atha, D. H, Wenz, H-M., Morehead, H., Tian, J., and O'Connell, C. D. Detection of p53 point mutations by single strand conformation polymorphism: analysis by capillary electrophoresis. *Electrophoresis* **19:**172–179, 1998.
25. Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., and Sekiya, T. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. USA* **86:**2766–2770, 1989.
26. Eng, C., Brody L. C., Wagner, T. M. U., et al. Interpreting epidemiological research: blinded comparison of methods used to estimate the prevalence of inherited mutations in BRCA1. *J. Med. Genet.* **38:**824–833, 2001.
27. Anderson, T. I. and Borresen, D. A.-L. Alterations of the TP53 gene as a potential prognostic marker in breast carcinomas. Advantages of using constant denaturant gel electrophoresis in mutation detection. *Diagn. Mol. Pathol.* **4:**203–211, 1995.
28. Keller, G., Hartmann, A., Mueller, J., and Hofler, H. Denaturing high pressure chromatography (DHPLC) for the analysis of somatic p53 mutations. *Lab. Invest.* **81:**1735–1737, 2001.
29. Ahrendt, S. A., Halachmi, S, Chow, J. T, et al. Rapid p53 sequence analysis in primary lung cancer using an oligonucleotide probe array. *Proc. Natl. Acad. Sci. USA* **96:**7382–7387, 1999.

21 Laboratory-Developed Tests in Molecular Diagnostics

ANDREA FERREIRA-GONZALEZ AND CARLETON T. GARRETT

1. INTRODUCTION

One of the fastest growing areas in clinical laboratory testing is molecular diagnostics. This growth has been in part the result of the tremendous amount of knowledge gained in the last decade from the Human Genome Project regarding organization, regulation, and expression of genomic information in humans and microbes. This knowledge has allowed us to better understand the pathogenesis of many diseases at the molecular level. As a consequence, diseases are being increasingly defined in terms of their molecular pathogenesis. This has lead to the development of new clinical molecular assays for diagnosis, prognosis, selection of therapeutic modalities, and monitoring of diseases. Until a few years ago, practical laboratory methods for detecting differences in nucleic acid sequences were not sufficiently simple or robust for the clinical laboratory. Recent advances in new technology, instrumentation, and efforts in standardization have overcome many of these limitations. The development and introduction of new technologies that allow automation of the testing process, as well as higher-throughput testing, have improved the diagnosis of diseases and patient care. In addition to the increasing number of kits and reagents used for molecular diagnostics approved by the Food and Drug Administration (FDA), a large number of molecular tests are still manufactured in-house by different clinical laboratories.

2. CLINICAL TESTS FORMATS

There are two major clinical assay formats widely used in molecular diagnostic laboratories. One of the formats are those tests developed by in vitro manufacturers and the other are assays developed by each laboratory (LDA), also known as home-brew assays.

2.1. TESTS DEVELOPED BY *IN VITRO* MANUFACTURERS This first category of tests is composed of kits developed by manufacturing companies to provide quality-controlled reagents for the performance of the entire molecular test for a determined clinical condition. For example, kits for monitoring patients diagnosed with human immunodeficiency virus-1 (HIV-1) infection through quantification of plasma HIV-1 RNA. These kits usually include the reagents necessary for nucleic acid isolation, amplification, and detection/quantitation. These complete kit assays usually include information about sensitivity, specificity, and tolerance limits for a particular clinical condition. They can be labeled by the manufacturer as FDA-approved, FDA-cleared, for research use only (RUO), or for investigational use only (IUO). Recently, we have seen the development and introduction of automated platforms for molecular diagnostic assays. A large number of these platforms have been developed by different in vitro diagnostic manufacturers, to provide automation for a single step or several steps of the testing procedure. Single-step automated instruments have been incorporated into molecular diagnostic laboratories, such as automated nucleic acid extractors, robotics for reagent preparation, and automated DNA sequencers. More recently, automated instruments that perform more than a single function have been developed and introduced in molecular diagnostic laboratories. An automated platform introduced by Roche Molecular Systems is the COBAS Amplicor® analyzer. This instrument contains a thermocycler for nucleic acid amplification, a 3D Cartesian arm with a gripper, heating blocks, wash stations, and even a colorimeter reader. It is anticipated that fully automated instruments will be developed and implemented in the near future. Analytical and clinical validation of these kits drastically differs depending on the labeling. Validation procedures will be discussed later.

2.2. LABORATORY DEVELOPED TESTS Laboratory developed tests are those tests (LDTs) fully developed and validated by the laboratory that performs them. They are also referred to as "home-brew" tests. Usually, these tests use a combination of reagents that are purchased separately from a variety of manufacturers. Each laboratory determines the performance characteristics of the assay for a specific clinical condition and a particular patient population. There are a number of different commercially available kits that are developed by a manufacturer to provide quality-controlled reagents to perform a particular step in molecular testing. For example, kits are commercially available for nucleic acid extraction, amplification

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

reagents including controls, in vitro amplified nucleic acid detection systems, and so forth. In some instances, the laboratory develops a particular molecular test by combining two or more kits from the same or even different manufacturers for a determined clinical condition in a particular patient population. The analytical and clinical validation of the entire testing process is the responsibility of each laboratory. Analytical and clinical validation of each step and the entire procedure will be discussed in a later section.

2.3. ANALYTE-SPECIFIC REAGENT RULE AND THE ROLE OF THE FDA IN THE DEVELOPMENT OF LDT A new term has entered the lexicon of molecular diagnostics. The term "analyte-specific reagent" (ASR) was devised by the FDA to refer to reagent(s) used in LDT that confers specificity for detecting a particular analyte. The FDA has been involved in the review of in vitro diagnostic devices (IVDs) since the introduction of the Medical Device Amendment of 1976. This amendment expanded the FDA jurisdiction to include all reagents, assays, and equipment used for the diagnosis and/or treatment of diseases. The Safe Medical Devices Act of 1990 further classified medical devices into three categories that require different levels of regulatory stringency. Class I devices are regulated as general controls, such as registration of the manufacturer with the FDA, listing of all medical devices and/or drugs manufactured by a company, manufacture devices and/or reagents under good manufacturing practices, appropriate record keeping and restrictions on the use, sale, and distribution of the device. Class I devices are subject to premarket notification submissions to establish "substantial equivalence" (510K review), even though the FDA can exempt a Class I device from premarket notification. Most Class I in vitro diagnostic devices are considered of low risk and are often exempted from premarket review. Class II devices require special controls in addition to general controls to assure the safety and effectiveness of the IVD. Special controls include performance standards, postmarket surveillance of product problems, and guidelines for product use. Class II devices are considered of moderate risk and subject to premarket notification submission requirements or 510K review. Class III devices are considered high-risk products and are regulated by premarket approval applications or PMA review. Furthermore, devices are classified by generic device types into six categories: (1) general-purpose reagents, which are Class I devices and mostly exempted from any review; (2) research use only (RUO), where FDA review is expected within 30 mo; (3) investigational use only (IUO), where the IVD is used to establish characteristics for FDA review; (4) kits for analysis (Class II or III requiring FDA review; (5) immunohistochemicals (Class I, II, or III); and, most recently, (6) analyte-specific reagents or ASRs. Most of the ASRs are considered Class I, but depending on their use, they could also be classified as Class II or Class III.

There are other mechanisms to offer testing to clinicians and health care providers, such as LDTs. The FDA considers LDTs as medical devices, and such products are subject to FDA regulatory oversight. However, with very few exceptions, the agency has decided not to execute its authority in this area. This decision was based not only on the limited resources available to the FDA but also on the understanding that laboratories developing LDTs for clinical use are regulated by Center for Medicare Services (CMS) under the provisions of CLIA'88. Even with this understanding, the FDA decided in 1998 to propose a modification to its regulations of oversight by introducing the new IVD category, named Analyte-Specific Reagent (ASR). The purpose of the FDA authorities was to introduce a mechanism by which they could clarify their role in the oversight of these devices and to use their regulatory authority to improve general controls and labeling applied to these devices without imposing a new burden for premarket review. In other words, under the ASR rule, the FDA has shifted the regulatory focus away from the assays developed in the laboratory in favor of regulating the reagents purchased by laboratories to develop LDTs.

As part of this new rule, the FDA defines ASRs as "antibodies, both monoclonal and polyclonal, specific receptor proteins, ligands, nucleic acid sequences, and similar reagents, which, through specific binding or chemical reactions with substances in a specimen, are intended for use in a diagnostic application for identification and quantification of an individual chemical substance or ligand in biological specimens." The FDA recognized ASR as the active ingredient of LDTs, that when used in combination with commonly used reagents and general laboratory instrumentations could be the basis for a LDT. In addition to defining ASR, the FDA proposed a series of controls and restrictions that apply to ASRs in order to ensure their quality and consistency and to clarify that laboratories that set up these tests are responsible for the test performance. Interestingly, these new controls applied to both manufacturers and laboratories developing LDTs that use ASRs. Another part of this rule mentions that manufacturers are required to register and list their ASR products with the FDA and follow good manufacturing practices. They are also expected to report adverse events resulting from product failure. Most of these products are exempted from premarket review but are subject to restricted labeling and use. The FDA also limited the labeling of these reagents to "description of the identity and purity including source and method of acquisition of the ASR in addition to standard information already required for general purpose reagents," with the specific disclaimer reading "Analyte specific reagent, analytic and performance characteristics are not established." Specifically, ASRs must be labeled to reflect that these reagents are a component used in LDTs and can only be sold to CLIA-certified high-complexity laboratories. Not every ASR that falls within this definition will be regulated under this rule. ASRs are excluded from the provisions of the rule when they are sold to IVD manufacturers or organizations that use reagents to develop tests for purposes other than providing diagnostic information to patients and practitioners (i.e., academic, research, forensic settings)

Moreover, laboratories developing LDTs are required to meet high-complexity testing requirements, to establish, maintain, and monitor the performance characteristics of the LDT, as specified in the CLIA regulations, and to label these LDTs with a specific statement. The FDA statement reads: "This test was developed and its performance characteristics determined by (Laboratory Name)." It has not been cleared or approved by the U.S. Food and Drug Administration." The rule became effective November 23, 1998. This clearly shifts the burden of

analysis of performance characteristics from the manufacturer to the user. Hence, the expectation under the ASR rule is that the laboratory will develop, evaluate, and document the efficacy of diagnostic tests using ASRs and will determine whether there is a need to invest more resources in the proper evaluation of the ASR performance in a laboratory setting.

The restrictions on ASRs are not intended to apply to products developed by laboratories for in-house use. The FDA's preface notes that "the focus of this rule is the classification and regulation of ASRs that move in commerce, not tests developed in-house by clinical laboratories or ASRs created in-house and used exclusively by that laboratory for testing services." In other words, the ASR rule does not state that a laboratory needs to have ASR primers and probes. If primers and probes are made commercially and sold to clinical laboratories, the company must comply with the ASR rule and label the products accordingly. These tests also need to be accompanied by a disclaimer regarding the lack of FDA approval.

3. IMPLEMENTATION OF LABORATORY DEVELOPED TESTS

3.1. TEST SELECTION Similar to any other clinical test, the primary goal of molecular tests should be to provide reliable and timely results necessary for patient care. The introduction of new molecular tests has the primary goal of improving patient care. Clinical needs, as well as advances in basic and translational research, often prompt the development of new molecular tests. Several key elements are highly important during test selection. It is necessary to realize that any new test should aid in cost-effective patient care and provide a less expensive or more effective diagnosis and/or patient management. It is also important not only to make test selection based solely in the cost of tests performance but also to determine how the new test would impact the overall care and management of the patient. There are several circumstances in which molecular tests seem to increase the cost of managing patients; however, the introduction of a new test can make this management more effective. One example of this situation was the introduction of HIV-1 viral load to clinical practice for monitoring disease progression and effectiveness of drug therapy in individuals infected with HIV-1. Current standard of practice for patients infected with HIV-1 is to treat these patients with a combination or cocktail of different drugs, including protease inhibitors, nucleoside inhibitors, and non-nucleoside reverse transcriptase inhibitors, all of which are very effective in reducing morbidity and mortality in a large number of HIV-1-infected patients. Even though a new test was added to the battery of tests used for these patients increasing the cost of their management, it has provided a tool that rapidly and accurately determines the effectiveness of an expensive drug treatment. Furthermore, mutations in the HIV-1 genome associated with resistance to antiretroviral therapy have been shown to be a major cause of treatment failure in patients infected with HIV-1. Despite the increased number of antiretroviral agents currently available, virologic failure remains a significant problem. Drug-resistance testing is designed to identify gene mutations that suggest reduced drug susceptibility. These resistance mutations can be assessed by a genotyping test that probes for specific mutations within the HIV-1 genome. The widespread use of these drugs and virologic benefits of resistance testing have prompted the development of formal guidelines by expert panels for clinical use of this test. HIV-1-resistance testing has resulted in better clinical outcomes and lower incidence of virologic failures, which are associated with lower cost of treatment. These examples show the impact of the introduction of new tests on patient's management; however, there are many examples of the cost-effectiveness of replacing existing laboratory tests. In addition, it is important to realize that for a molecular test to be effective, it should provide increased sensitivity, increased specificity, or reduced turnaround time that directly impacts on the patient's management.

The identification of tests that would be useful at each medical center is primarily the responsibility of the director of the molecular diagnostic laboratory. Sometimes, this could be aided by the performance of a review of the send-out list of tests for the medical center. Sometimes, key players, clinicians and pathologists, could be a valuable source for identifying the needs for each specific medical center. It is important for the laboratory directors to clearly identify an area where there is a perceived need for improving the tools available for patient diagnosis and management. It is important that during the process of test selection, realistic matching of the technical capabilities within the laboratory with the real-world clinical needs in terms of test volume, required turnaround time, and associated costs of performing the assay are matched. It is also important to estimate the cost of development and test volume, this last parameter is critical to be considered, not only to assure recovery of test development but also to maintain proficiency by the testing personnel, avoiding excessive reagent outdates. The establishment of a period of clinical trial for evaluation of the clinical utility of a particular test for a particular clinical condition is an important market approach. If carefully designed, this approach allows the laboratory to directly work with the end user of the test and provide an avenue for these individuals to understand the clinical utility and its limitations and eventually more effective utilization of the particular test. Laboratory developed assays are costly to develop and validate, and for that reason, it is recommended to be sure that FDA review for a similar assay is not pending from a commercial manufacturer for the same analyte.

Once a particular analyte and clinical condition have been identified, it is important to give some consideration to the different methodologies that might be available for evaluating the particular analyte or disorder. Furthermore, careful consideration of the advantages and disadvantages of the different methodologies and the particular condition that is under evaluation should be performed. As an example, we can mention the detection of t(9;22) in patients with chronic myelogenous leukemia (CML). Conventional cytogenetic analysis is the recommended test for detecting t(9;22) in newly diagnosed patients; however, these tests require viable cells or at least 10% of blasts in peripheral blood to obtain reliable results. Cytogenetic analysis carries a sensitivity of 5%. Fluorescent in situ hybridization (FISH) allows the detection of the translocation in metaphase or interphase cells and can be applied directly to blood leukocytes or other non-dividing cells with a

Table 1
Conducting the Clinical Testing Process

Activity	*Considerations*
Reagent preparation	1. Perform in cleanest environment. 2. Store working stock solutions in single-use aliquots. 3. Quality control each new set of reagents prior to use in clinical testing. Use sample with low copy number for sensitivity evaluation. 4. Preparation of master mixes to reduce variability and errors. Established shelf life and storage conditions for short-term and/or long-term storage
Specimen collection	1. Establish acceptable tolerance limits for each specimen type to be tested (storage temperature, transport time, anticoagulant, etc.). 2. Distribute protocols for proper specimen handling to all potential users. 3. Capture clinical and analytical information on requisition.
Specimen processing	1. Specimen must be received and stored in preamplification (clean) laboratory. 2. Develop guidelines against specimen mix-up and to preserve integrity of target sequence.
Analysis of specimen	
(a) Extraction procedure	1. Evaluate extraction procedures for presence of inhibitors and factors that decrease yield of target. 2. Internal control added to sample at the time of extraction to determine false negatives because of inhibitors or determine rate of inhibitors by some other means. If an internal control is not supposed to be evaluated with each patient's sample, the rate of false negatives should be stated on the report in a disclaimer for negative results.
(b) Assay setup, amplification	1. Optimize concentration of primers, $MgCl_2$, dNTPs; volume; cycling conditions and detection and detection system. 2. Develop guidelines to minimize possibility of contamination by template nucleic acid or amplicon (see sections on quality control). 3. Controls must be processed in the same manner than patient specimens. 4. Develop guidelines for setting up assays to avoid cross-contamination of specimens and controls.
Interpretation and report	1. Develop guidelines for interpretation and report. 2. Interpretation should be performed by at least two individuals independently. 3. Develop guidelines for report distribution.

sensitivity of 0.5%. Another method that can be used for the detection of t(9;22) is Southern hybridization analysis, which reliably identifies the BCR gene rearrangement using probes to the major or minor breakpoint with a sensitivity of about 5%. The reverse transcriptase–polymerase chain reaction (RT-PCR) is the most sensitive method for the detection of t(9;22). Amplification methods are capable of detecting 1 tumor cell among 100,000 normal cells, which makes this test suitable for detection of minimal residual disease. Real-time RT-PCR techniques are now available for a precise quantification of the chimeric transcripts, which has some prognostic value. Before embarking in the development of a particular methodology, it is important to keep in mind the specific disease and the advantages and disadvantages of each procedure for diagnosing that particular clinical condition.

3.2. PATENT ISSUES A large number of in vitro amplification procedures are patented and a license agreement for their use must be obtained; if not, the laboratory using the procedure is liable to prosecution for patent infringement. Traditionally, licensure is obtained for a particular procedure by purchasing an FDA-approved set of reagents sold by a manufacturer that holds the patent or an exclusive license to the amplification process. This approach is very limited for molecular diagnostic tests because there are few tests approved by the FDA. Thus, a laboratory using a particular patented process for clinical use must first negotiate a license agreement with the patent holder.

Almost identical considerations must be given to the use of sequence information required to design an assay. Sequences of newly discovered genes (e.g., human genes and virus genomes) are frequently patented by their discoverers, and the user of the sequence information without a license agreement runs the same risk of liability for patent infringement. Unfortunately, with the current environment and diversity of sources for sequence information, it is not always clear whether the sequence has been patented or who owns the patent. For this reason, before undertaking the development of any test on published sequences where a major commitment of resources is planned, it is advisable to check with the investigators that first described the sequence to determine patent issues. Also, careful review of the patent database and pending patent databases could be very useful (United States Patent and Trademark Office home page: www.uspto.gov).

3.3. ASSAY DESIGN AND DEVELOPMENT Once specific analytes, assay techniques, and specimen types have been identified, the assay design and development can begin. Table 1 describes the different steps of the testing process that need to be taken into consideration when designing an LDT. The first step in introducing an LDT is to optimize each step of the analytical process, which includes nucleic acid extraction, amplification, detection, quantification, and result interpretation. Several review and research articles have provided detailed descriptions of the key parameters that might influence the

Table 2
Guidelines and Standards for Molecular Diagnostics Testing

Organization	*Guideline or standard*	*Address*
NCCLS	MM1-A Molecular Diagnostic Methods for Genetic Diseases MM2-A Immunoglobulin and T-Cell Receptor Gene Rearrangement Assays MM3-A Molecular Diagnostic Methods for Infectious Diseases MM-5-A Nucleic Acid Amplification Assays for Molecular Hematology MM-6-A Quantitative Molecular Diagnostics for Infectious Diseases MM-7-P Fluorescence in Situ Hybridization Methods for Medical Genetics MM-8-P Measurement and Interpretation of Trinucleotide Repeats	Wayne, PA; www.nccls.org
ACMG	Standards and guidelines for clinical genetic laboratories: Policy Statements Prenatal Interphase Fluorescence In Situ Hybridization ACMG Position Statement on Multiple Marker Screening in Women 35 and Older Fragile X Syndrome: Diagnostic and Carrier Testing Technical standard and guidelines for Fragile X: the first of a series of disease Laboratory standard and guidelines for population-based cystic fibrosis carrier screening (*Genet Med* **3**:149–154, 2001) Factor V Leiden Working Group–American College of Medical Genetics consensus statement on Factor V Leiden mutation testing (*Genet Med* **3**:139–148, 2001). Statement on Storage and Use of Genetic Materials Statement on Multiple Marker Screening in Pregnant Women Statement on Use of Apolipoprotein E Testing for Alzheimer Disease. Points to Consider: Ethical, Legal, and Psychosocial Implications of Genetic Testing in Children and Adolescents Diagnostic Testing for Prader–Willi and Angelman Syndromes: Statement on Population Screening for BRCA-1 Mutation in Ashkenazi Jewish Women Principles of Screening: Report of The Subcommittee on Screening of the American College of Medical Genetics Clinical Practice Committee Position Statement on Carrier Testing for Canavan Disease Statement on Genetic Testing for Cystic Fibrosis Administrative office, 9650 Rockville Pike, Bethesda. MD 20814–3998	ABMG/ABGC/ACMG; II Administrative office: 9650 Rockville Pike, Bethesda, MD, 20814–3998
ASHI	Standards for Molecular Histocompatibility and Immunogenetic Testing	ASHI, P.O. Box 15804, Lenexa, KS 66285–5804
NIH-DOE	Task Force on Genetic Testing-Promoting Safe and Effective Genetic testing	www.nhgri.nih.gov/Policyandpublicaffairs/Elsi/tfgentest
FDA	Guidance for industry in the manufacture and clinical evaluation of *in vitro* tests to detect *in vitro* nucleic acid sequences of HIV-1-Draft Guidance Guidance for industry and/or FDA reviewers staff—Premarket approval applications for assays pertaining to Hepatitis C virus (HCV) that are indicated for diagnosis or monitoring of HCV infection or associated disease-Draft Guidance	www.fda.gov/cber/gdlns/nashiv.pdf; www.fda.gov/cdrh/ode/1353pdf
AMP	Recommendations for in-house development and operation of molecular diagnostic tests	www.amp.org
Technical Working Group on DNA Analysis Methods	Guidelines for a Quality Assurance Program for DNA Analysis	*Crime Lab. Dig.* **18**:44–75, 1991.

performance of different *in vitro* nucleic acid methodologies, including standard PCR, either uniplex or multiplex. Optimization will not be discussed in this chapter and the reader is recommended to review these citations for further information. When optimization is carried out for each step separately, it is important to realize that when all steps are placed together, reoptimization is required in most of the cases. After optimization of the assay, it is necessary to evaluate and document preanaytical variables that might have an impact on the performance characteristics of the assay. Some of the most common preanalytical variables are specimen type, transport, storage, and handling requirements, as well as interfering substances such us lipids, hemoglobin, billirubin, and so forth.

The use of internal controls to detect the presence of inhibitors or nucleic acid degradation is extremely important. One of the major advantages of using internal controls is that they can mimic the nucleic acid present in a patient specimen and be subject to the entire testing protocol from nucleic acid extraction to detection. A common approach to development of internal controls is the creation of synthetic materials such as in vitro synthesized plasmids that would contain a modified target sequence, which could be spiked into the specimen before

Table 3
Checklist for Verification of Laboratory Developed Tests

Name of test	State the name of the test, including the trade or proprietary information if exists; be sure that the name identifies the particular disease or condition or gene locus to be analyzed
Intended use	What the test measures and for what purpose; identify the particular parameter(s) that the test measures, point mutation/deletion/insertion, and indicate the use of the test (e.g., diagnosis, prenatal, carrier status, screening, presymptomatic, etc)
Indications for use	Provide clinical condition(s) CF. Use reference standard definitions as found in OMIM.
Method category	Identify test method
Testing procedure	Information with regard to specimens, specimen handling, analyte extraction/isolation, analyte storage, description of the test procedure, data reports, expected results, technical interpretation of results
Test results	Representative examples of results
Analytical validity	Analytical sensitivity, analytical specificity, precision, etc.
Quality control and quality assurance	Delineate the QC and QA program
Assay limitations	Briefly delineate and discuss potential limitations
Clinical data	Primary objective of the study, clinical condition evaluated, patient population demographics, sample size estimate
Clinical validity	Clinical sensitivity, clinical specificity
Reporting of tests results (clinical interpretation)	Clinical interpretation
Clinical utility	Potential clinical benefit to patient and physician

testing. The most common internal controls are synthetic segments of nucleic acid that use the same primer sequence as the target molecule, with an internal portion of the sequence that is unique to the internal control, as to allow separate detection from the target and quantitative standard with the use of specific probes for each species. In addition, this internal control can be used as an internal calibrator. An internal calibrator contains a predetermined amount of the modified target that can be added directly to a clinical specimen and undergo the same manipulation as the target of the patient specimen. One caveat for internal controls is that they must have the same or a very similar efficiency of amplification. Moreover, when the internal control fails to be detected, it is not possible to determine the specific reason (i.e., inhibitors or other amplification problems). In addition, it is important to use a low amount of the internal control to avoid competition with the target of the patient specimen. When detecting and/or quantifying RNA from tissue or cells, amplification of a housekeeping gene can be used as an internal control. Most of the housekeeping genes are highly abundant and might not be the most appropriate internal control. The use of RNA from a gene with approximate abundance as the target sequence is preferred.

4. TEST VALIDATION

As with any other area of the clinical laboratory, the introduction of a new test requires proper validation. There are major differences between the implementation of a FDA-approved test compared to one that is not. CLIA'88 provides specific provisions for validation of these assays. Laboratories implementing FDA-approved tests need to verify and document the performance characteristics of the test for the indications for use in populations similar to those in which the manufacturer has established performance. On the other hand, implementation of in-house-developed assays requires a more involved process with extensive documentation on test performance, in addition to quality control programs to assure daily performance of the test. Table 2 provides a list of guidelines and standards for molecular diagnostics testing that can be used as a reference during assay development and verification. In addition, Table 3 describes the different aspects of documentation that would be useful to capture during this process.

Verification of a LDA is a complex process that can be divided in two phases: analytical and clinical verification. These parameters provide information about the analytical performance of the test. The other phase of the verification process is the clinical verification. This phase provides information about the clinical utility of the test with regard to the intended use. Determination of the intended use provides information about the appropriate settings, including disease states and populations for which the test can be useful. It is important to point out that "intended use" refers to the DNA or RNA that a molecular assay detects or quantifies, whereas "indications for use" refers to the clinical syndrome/condition for which the assay can be used for diagnosis, management or monitoring of disease.

5. ANALYTICAL VERIFICATION

Before a new or improved LDT is introduced into the laboratory menu, careful evaluation of performance characteristics of the assay under laboratory conditions needs to be done. In addition to evaluating these characteristics, undertaking analytical verification can provide useful information with regard to issues of practicality that will have to be addressed before introducing the LDA. It is important to point out that the performance of analytical validation programs has been challenging because of the lack of standards for a large number of nucleic acid targets. This shortfall has impacted the laboratory's ability to determine assay sensitivity and accuracy. Table 4 describes a number of different suppliers of commercially available reference materials that can be used during analytical verification and as a source of material for quality control purposes.

Analytical verification provides information about the performance characteristics of the assay. This section provides information about the design and execution of analytical verification studies. As part of the analytical verification, laboratories must determine the assay's analytical sensitivity, analytical

Table 4
Commercial Sources for Control Materials

Company	*Available material*
Boston Biomedica Inc. (www.bbii.com)	Quantitative and qualitative controls, panels for HCV, HIV, HBV, CMV, *Chlamidya trachomatis*, *Mycobacterium tuberculosis*
Advanced Biotechnology Inc. (www.bioresearchonline.com/storefronts/abi.html)	Control DNA, virus and cell lines (HIV, HTLV, EBV, CMV, HSV, VZV, HCV, SIV, rubella, native and recombinant products purified viruses, antigens, proteins, antibodies, research kits)
National Institute of Standards and Technology (www.nist.gov)	PCR-Based DNA profiling standard for human identity testing
Coriell Institute for Medical Research (arginine.umdnj.edu)	The Coriell Cell Repositories provide essential research reagents to the scientific community by establishing, verifying, maintaining, and distributing cells, cultures, and DNA derived from cell cultures for inherited disorders, microsatellite fingerprinting
Stratagene (www.stratagene.com)	Universal Human Reference RNA for microarray technology
AcroMetrix (www.acrometrix.com)	HIV, HCV, and HBV
Promega Corporation (www.promega.com)	A human mitochondrial DNA standard reference material for quality control in forensic identification, medical diagnosis and mutation detection
American Type Culture Collection	Bacteria, bacteriophages, cell lines and hybridomas, filamentous fungi and yeast, tissue cultures, viruses, etc.

specificity, accuracy, and precision. In addition, for quantitative assays, the information regarding the linear dynamic range or reportable range of the overall process provides valuable information to determine when a measurement or change in the quantity of the analyte is considered clinically relevant or the result of inherent test error. A number of national and international organizations are taking steps in developing standard reference materials. The National Institute of Standards and Technology (NIST) developed one of the first nucleic acid standard reference materials for human identity testing. More recently, the World Health Organization (WHO) introduced a standard reference material for hepatitis C that has been used for verification of nucleic acid tests (NATs) for screening of blood and blood products and has also provided useful material for other in vitro diagnostics devices. There are currently three registered WHO standards for virologic molecular testing. The WHO HCV (heptatitis C virus) standard (HCV 96/790) consists of a lyophilized material, which contains 50,000 units per 0.5 mL or 100,000IU/mL. This material was assigned a value of 5.14 log genome equivalents/mL when tested by a variety of different molecular methods in different laboratories. Following the HCV standard, WHO introduced an HIV standard, HIV 97/656. Similar to the HCV standard, the HIV standard consist of a lyophilized preparation of HIV-1-genotype B virus-positive plasmapheresis donation. Several laboratories have tested this material with different technologies, obtaining a mean value of 4.79 log genome equivalents/mL, which was assigned a value of 100,000 IU/mL. In addition, the HBV (hepatitis B virus) standard 97/746 has been assigned a value of 1,000,000 IU/mL. Reference panels calibrated to WHO standard reference material are commercially available (Boston Biomedica Inc., Boston, MA).

Because of the lack of standards, laboratories have relied in the use of reference materials of different kinds to validate different tests. Laboratories can develop their own reference material for analytical validation, which can be used subsequently for monitoring daily performance of the assay. Creating an in-house reference material involves the use of an independently established method for determining target nucleic acid concentration. Alternatively, the laboratory can conduct studies to compare with another established assay. For example, samples can be split for a comparison study with another laboratory that performs a similar molecular test. These samples might be available in-house or from an outside source such as a collaborating laboratory, government (CDC, FDA, or National Institutes of Health), or even a commercial supplier (Boston Biomedical, Accrometrix). Natural analytical reference materials are those that consist of a known analyte or known quantity of an analyte as it occurs naturally in the test matrix, purified from the test matrix, and derived by culture or even cell lines considered nonsynthetic reference materials. Some examples used by a number of laboratories are intact virus particles, bacteria naturally containing the target in their genome, cell lines containing a specific genetic change, plasmids, intracellular RNA or DNA, and so forth. Furthermore, in some instances, it has been very difficult to obtain a natural analytical reference material and laboratories have depended on the development of synthetic reference materials. These synthetic reference materials could be in the form of DNA, either single or double stranded, or RNA manufactured in vitro, which can be accurately quantified by several physical and/or biochemical methods. Such synthetic reference materials can include synthesized DNA in the form of oligonucleotides, single-stranded DNA produced by cloning recombinant phage, cloning into vectors such us plasmids, or a DNA fragment produced by chemical or physical method from a larger DNA molecule followed by purification. On the other hand, synthetic RNA reference materials can be generated by in vitro transcription of DNA templates.

5.1. PRECISION Every laboratory's measurement has an inherent error or random variation. The evaluation of the precision of an assay allows determining what constitutes a change in the analyte as a result of changes in clinical condition vs expected fluctuations of the laboratory measurement. Precision refers to the agreement of values obtained between replicates of

the same material. Evaluation of the precision should be measured for the entire process from nucleic acid extraction, amplification detection, and/or quantitation of the intended analyte. Precision studies should be carried out using test material or reference material that is similar or closely resembles the intended patient specimen. This could be achieved by using a serial dilution of a target positive specimen into a negative specimen. If a characterized patient specimen is not available, a reference material could be created by mixing a cell line, purified virus micro-organism, and so forth into a pool of patient's specimen known to be negative for the specific analyte. In addition, enough material should be prepared in order to last the entire experiment. This material could well be used for daily monitoring of the assay's performance. The concentration of the analyte spiked into the patient's specimen will depend on the type of assay being developed. For qualitative assays, a single reference control with a concentration close to the limit of detection is recommended. On the other hand, for quantitative methods, at least two concentrations should be tested. These two concentrations should span the linear dynamic range of the assays, and when possible, be close to the value used for clinical decision-making. Whenever possible, precision studies should be performed with more than one lot or batch of reagents and/or materials.

5.2. ACCURACY Accuracy of a method refers to its ability to measure the true value of the particular analyte. Determination of the accuracy of molecular methods has been challenging because molecular methods have proven to be more sensitive than a large number of well-established gold standard methods. Generally, when evaluating a new assay, the results obtained should be compared to the results obtained from an established assay, which is considered a "gold standard method." In the absence of a gold standard to compare the results to, a laboratory could purchase reference material to be used for the analytical verification.

5.3. QUANTITATIVE METHODS: DETERMINATION OF LINEARITY AND REPORTABLE RANGE Linearity of an assay is the measure of the degree to which a curve approximates a straight line. Furthermore, the linear range of an assay is the span of analyte concentration for which the final value output is directly proportional to the analyte concentration, with acceptable accuracy and precision. The boundaries of the linear range constitute the upper and lower limit of quantitation, but they are not necessary for the limit of detection. Determination of the linearity of a quantitative assay might be performed by testing at least four different levels of the analyte. Again, test material could be prepared by spiking the analyte into negative patient sample or by performing serial dilution of a patient specimen known to contain a very high level of the analyte.

5.4. ANALYTICAL SENSITIVITY Analytical sensitivity is the lowest amount of a specific analyte that the method can reproducibly detect. The analytical sensitivity represents the ability of a test to obtain a positive result in concordance with positive results obtained by a reference method. For quantitative molecular methods, the lowest amount that the method can detect might be different from the lower limit of quantification for a particular nucleic acid or micro-organism. The lower limit of quantification is the lowest amount of a nucleic acid sequence that can be detected with acceptable precision. Analytical sensitivity could be determined by performing serial dilutions of an appropriate number of samples containing different concentrations of the analyte.

5.5. ANALYTICAL SPECIFICITY Analytical specificity is the ability of an analytical method to detect and/or quantify what the analyte is intended to measure. One aspect of specificity that can easily be measured is the lack of crossreactivity with closely related nucleic acid sequences, organisms, and so forth. In addition, for infectious disease testing, it is important to determine lack of crossreactivity with nucleic acids from organism present in the normal flora or that would normally be present in a patient specimen.

In addition to the parameters already discussed, it is important to determine the effect that interfering substances might have on the ability of a test to detect and/or quantitate the anlyte of interest. An interfering substance is a component present in the patient specimens that interferes with the accurate detection and/or quantitation of a specific target. The source of the interfering substance could be endogenous or exogenous. Exogenous interfering substances could be drugs, parenteral nutrition, anticoagulants, and so forth. Furthermore, it has been shown that the use of some anticoagulants such us heparin could interfere with the amplification process. On the other hand, endogenous interfering substances could be result of pathologic conditions, (e.g., lipids, bilirubin, etc.). Several approaches could be used to address the effects of different interfering substances. The addition of nucleic acid targets, either purified nucleic acid, cells, or micro-organisms to a variety of different specimens that contain interfering substances can be a means to address this issue. Furthermore, the specific target could be added to specimens from patients with specific conditions or receiving specific drug treatment.

5.6. PREANALYTICAL CONSIDERATION Appropriate specimen handling, including specimen collection and transport conditions, is critical to the testing process to ensure specimen integrity. Inappropriate specimen handling could result in nucleic acid degradation, which can lead to false negatives and/or inaccurate quantitation of nucleic acid. This is critical for assays that detect and/or measure eukaryote RNA and viral RNA. The best specimen type and quantity should be determined because molecular tests have been applied to a variety of specimen types. Appropriate selection of specimen type will depend on a variety of factors, including the condition being assayed and the type of nucleic acid required for the test. The format of the molecular assay being developed could greatly affect the amount of specimen required; for example, tests that require samples to be run in duplicate would require more specimen. Specimen transport and storage should be evaluated for every assay and type of nucleic acid. Specimen transport and storage conditions could vary significantly between specimen type, analyte (RNA vs DNA), cells, and micro-organism and must be determined by each laboratory. Special determinations of specimen transport and storage is crucial for RNA, because it is highly susceptible to degradation by ubiquitous enzymes. Transport and storage conditions can vary greatly for different specimens, analyte, and assay type. These conditions could vary from specimen storage at room temperature to having to centrifuge samples, remove plasma or serum, and store them at –80°C until tested.

Spiking target-negative specimens with a known amount of purified target, micro-organism, or cells can be used to perform the assessment of the preanalytical variables mentioned earlier. The same approach can also be used to assess the effect on performance by lipids, hemoglobin, bilirubin, therapeutic drugs, and specimen anticoagulants.

6. CLINICAL VERIFICATION

Laboratory methods provide information for use in managing patients and addressing relevant clinical questions. The usefulness of a method depends on both the analytical performance and the clinical characteristics of the method (clinical sensitivity, specificity, and predictive value). These last factors are characteristic of the clinical application and not properties of the test. The clinical significance of a test should be defined in terms of the disease or syndrome that is the subject of the test. These include disease prevalence, possible outcome, and cost to the patients and others for incorrect information (false-positive and/or false-negative results). Validation of clinical utility of an assay is also a complex process that possesses many challenges. Clinical utility is refered to here as the indications for use. For example, the assay is intended to quantify a nucleic acid (analyte) but is indicated for a certain clinical condition or scenario. The first step of the analytical verification is to formulate a clinical question and to identify the target population to which the test will be applied. There are three integral parts in formulating a clinical question: (1) defining the subject population, (2) defining the management decision that needs to be made, and (3) to identify the role of the method in decision-making. When the clinical use of each assay is being considered, it is assumed that the assay analytical performance characteristics are very well understood. Known analytical limitations (e.g., reportable range for quantifying, precision for nonclinical samples) should be taken into consideration when determining the clinical use for each particular assay. Clinical verification requires evaluation of the clinical sensitivity of the test by evaluating an appropriate number of samples from patients who have been diagnosed with the disease. The population and type of sample should be fully described because the results for a method in a given population using a specified sample type might not be suitable for another population, or even for another sample type within the same population. It is important to define the purpose indications for use of the method prior to starting the clinical verification. For example, indication for use of an assay could be diagnostic, population screening, or confirmatory. Clinical specificity can be determined by analyzing samples from patients diagnosed with a different disease that might be confused with the indicated disease and appears in the differential diagnosis. In addition, specimens from healthy donors can be appropriate for determining the clinical specificity of the test. The other parameters that need to be determined are the positive and negative predictive values, as well as a predictive value of the test over the course of disease or therapy. It is of general practice for a clinical laboratory to determine sensitivity and specificity of a new assay by comparing the new assay with a gold standard method or a method currently used in the laboratory. In some instances, molecular tests have appeared to be more sensitive or specific than current gold standard tests. Resolution of discrepancies between the new and currently used or gold standard methods can pose some challenges. In a number of instances, using a different molecular method has resolved discrepancies. Furthermore, in order to determine the clinical utility of a new LDA, variables that could affect results and their interpretation should also be considered. Some of these variables could be assessed for infectious disease testing, like microbial host interactions, microbial dynamics, variants and mutations, replicative fitness of the microbial agent, and so forth.

It is important to point out that data generated during the clinical verification of each assay should be gathered and kept organized to satisfy the requirements by the different regulatory bodies. These user-generated data should be filed and maintained according to the certifier's criteria. Within the documentation, the laboratory should have available a list of indications for which the assay has been verified (Table 3).

7. SPECIAL CONSTRAINTS RELEVANT TO GENETIC TESTING

Genetic tests are performed for diagnostic and/or prognostic purposes, prenatal diagnosis, newborn screening, determination of carrier status, and presymptomatic and predictive testing. A number of limitations must be considered when offering genetic testing. The first one is that a test might not detect all the possible mutations that could be present in particular gene that causes a disease. A single gene can have a host of different mutations. A negative result cannot completely assure that the person tested does not carry a mutation in a particular gene. At the same time, a positive test result might pose different risks for different people. In addition, another limitation of genetic testing is related to the complexity of how certain genetic diseases developed. This poses a challenge mostly for genetic testing for predictive purposes, because predictive testing cannot provide certain answers for everyone that might be at risk for a disease.

8. QUALITY CONTROL AND QUALITY ASSURANCE OF THE TESTING PROCESS

Every molecular diagnostics laboratory should develop a comprehensive written quality assurance program. The objective of the quality assurance program is to objectively and systematically monitor and evaluate the quality and appropriateness of the test results. The quality assurance program should address every aspect of the testing process: preanalytical, analytical, and postanalytical processes. The program must include written policies and documentation for the education and training of personnel, continuing medical education, proficiency testing, internal and external inspections, including documentation of corrective actions for deficiencies cited, quality control programs for the clinical testing, equipment performance, and safety. Establishment of molecular diagnostic tests, particularly amplification assays, requires many considerations at every stage, including reagent preparation, specimen collection, specimen aliquoting, and performance of the actual assay. There are a number of considerations that apply to all molecular assays and that important to guarantee reliable results. Quality assurance programs must be developed to address the testing process during all phases of the testing process: preanalytical, analytical, and

postanalytical. The amount of quality assurance and quality control performed for any molecular assay will depend on the type of test employed, FDA cleared, "for research use only," off-label, or ASR. The preanalytic and postanalytic phases for FDA-cleared and research use only tests are very similar. As part of the preanalytical phase, special attention should be given to the specimen collection, transport, and storage. This is critical because in many situations, the results reported could influence treatment decisions for a particular patient; this decision should be reflective of the patient status and not of specimen handling or mishandling. Quality control of reagents is another critical aspect that needs to be considered for all laboratory programs. Laboratories should determine critical reagents for the overall testing procedure and these reagents should be tested for composition, concentration, purity, and functionality. All new reagents should be tested in parallel using new reagents and should be compared to those generated using current reagents in crossover studies and proven to be comparable before the new reagents are used in assays with clinical samples.

As with any laboratory tests, a well-thought-out and well-written laboratory procedure is a key factor for the reproducibility of the assay. It is one of the most important aids during hands-on training of new personnel. The procedure protocol should be written according specific guidelines set up by the National Committee for Clinical Laboratory Standards. Performance of nucleic-acid based methods usually requires specific setup and/or workflow.

A careful selection of controls is vital for the interpretation of results. Several types of control are used through out the execution of the assays in order to assure the appropriate performance of a specific assay. Negative and positive controls are required by CLIA'88 regulations and must be processed in every clinical test. Failure to obtain the correct result for any of the controls invalidates the entire test and requires retesting of all samples. Whenever possible, positive and negative controls should resemble a patient's specimen as much as possible. Furthermore, a positive control should represent a clinically relevant range of the nucleic acid target sequence in a background of a negative nucleic acid target sequence. The negative control should represent a negative nucleic acid target. In addition to the negative and positive controls, a blank control should be included with every assay, containing all of the components of the reaction mixture but nucleic acid. In addition to these controls, in some instances it is imperative to add an internal positive control. There are many circumstances in which the presence or absence of an amplicon makes the diagnosis. In these circumstances, when a negative result is obtained, it is not clear if the absence of such amplicon is the result of the absence of the target sequence in the patient specimen or, for example, the result of the presence of inhibitors. In order to avoid this situation, it is recommended to perform the amplification of an internal positive control. The addition of sensitivity controls in which the last control is not detected is very valuable for monitoring not only the performance of the assay over time but also for monitoring the presence of amplicon contamination.

It is extremely important for LDTs to implement a quality control program for validating the robustness, purity, and performance of every critical reagent of the testing process. Every critical reagent should be tested before being approved for clinical testing. Tolerance limits for every critical reagent should be established. Whenever possible, tolerance limits should be established using a quantitative measurement in order to avoid subjective evaluation of the quality of the critical reagent.

BIBLIOGRAPHY

1. Association for Molecular Pathology statement. Recommendations for in–house development and operation of molecular diagnostic tests. *Am. J. Clin. Pathol.* **111(4):**449–463,1999.
2. Carter, M. A. Ethical aspects of genetic testing. *Biol. Res. Nurs.* **3(1):** 24–32, 2001.
3. Dequeker, E., Ramsden, S., Grody, W. W., Stenzel, T. T., and Barton, D. E. Quality control in molecular genetic testing. *Nat. Rev. Genet.* **2(9):**717–723, 2001.
4. Ferreira–Gonzalez, A., Wilkinson, D. S., and Garrett, C. T. Establishing a molecular diagnostics laboratory, in *Management Clinical Laboratory Methods* in 20th ed. John B Henry (ed.).
5. Food and Drug Administration Modernization Act. Public Law 105–115, 111 Statute 2296, 1997.
6. Kaul, K., Leonard, D. G. B., Gonzalez, A., and Garrett, C. T. Oversight of genetic testing: an update. *J. Mol. Diagn.* **3(3)**:85–91, 2001.
7. McGovern, M. M., Benach, M. O., Wallenstein, S., Desnick, R. J., and Keenlyside, R. Quality assurance in molecular genetic testing laboratories. *JAMA* **281(9):**835–840, 1999.
8. Medical Device Amendments of 1976. Public Law 94–295, 90 Statute 539, 1976.
9. Medical Devices: classification/reclassification; restricted devices; analyte specific reagents. Fed. Reg. **62:**62,243–62,260, 1997.
10. MM3—A Molecular Diagnostics method for infectious diseases; Approved guidelines, 1995.
11. Mulcahy, G. M. The integration of molecular diagnostic methods into the clinical laboratory. *Ann. Clin. Lab. Sci.* **29(1):**43–54, 1999.
12. Niesters, H. G. Quantitation of viral load using real-time amplification techniques. *Methods* **25(4):**419–429, 2001.
13. Niesters, H. G. Clinical virology in real time. *J. Clin. Virol*. **25 (Suppl. 3):**3–12, 2002.
14. Neumaier, M., Braun, A., and Wagener, C. Fundamentals of quality assessment of molecular amplification methods in clinical diagnostics. International Federation of Clinical Chemistry Scientific Division Committee on Molecular Biology Techniques. *Clin. Chem.* **44(1):**12–26, 1998.
15. Prence, E. M. A practical guide for the validation of genetic tests. *Genet. Test.* **3(2):**201–205, 1999.
16. Safe Medical Device Act of 1990. Public Law 101–629, 104 Statute 4523, 1990.
17. Schwartz, M. K. Genetic testing and the clinical laboratory improvement amendments of 1988: present and future. *Clin. Chem.* **45(5):**739–745, 1999.
18. Swanson, P. E. Labels, disclaimers, and rules (Oh my!). Analyte-specific reagents and practice of immunohistochemistry. *Am. J. Clin. Pathol.* **111:**445–448, 1999.

VI Applications of Molecular Diagnostics for Genetic Diseases

22 An Overview of Molecular Genetics

ELAINE WEIDENHAMMER AND GREGORY J. TSONGALIS

1. INTRODUCTION

Sir Archibald Garrod concluded in the early 1900s that predisposition to disease is dependent on each individual's chemical composition *(1)*. Beadle later described the one gene–one enzyme concept, which emphasized that biochemical processes are genetically controlled and that mutations in any given gene would result in a defective biochemical reaction *(2)*. In 1953, Watson and Crick described for the first time the structure of the DNA molecule, now recognized as the "blueprint" of all living things *(3)*. In their report, they described in detail the nature of complementary base-pairing as part of the stoichiometry necessary for the double-stranded DNA structure to maintain its integrity. The current concepts of molecular mechanisms of disease have evolved from these early observations. Our ability to detect both intrachromosomal and extrachromosomal genetic alterations at the molecular level has led to a revolution in laboratory medicine, by promoting an understanding of molecular pathology (i.e., the molecular mechanisms of disease processes). The fact that many human diseases can now be associated with defects at the gene level has led to an understanding of the mechanisms of disease inheritance *(4–8)*. Molecular genetics provides an avenue for dissecting complex pathophysiological processes into specific gene defects. The functions and interactions of these genes has been termed "genomics" *(9)*.

2. THE HUMAN GENOME

Biological information exists as three fundamental forms: DNA, RNA, and protein. Each is accompanied by a wide degree of structural and functional complexity. Resulting in part from the major strides made by researchers in the development of recombinant DNA technologies and from the unprecedented successes of the Human Genome Project, we are now capable of providing direct and indirect molecular diagnostic assays for many human diseases at the level of the nucleic acid. The human genome is thought to comprise 30,000–40,000 genes neatly packaged into a total of 46 chromosomes: 22 pairs of autosomes and 1 pair of sex chromosomes (X or Y) (Fig. 1). These 46 chromosomes represent the diploid number, found in most cells in the body. A haploid genome, found in sperm and egg cells (gametes), contains half of the total genetic material, or 23 chromosomes. The haploid genome contains approx 3×10^9 basepairs (bp); on average, then, each chromosome contains 1,000–1,500 genes dispersed within 1.3×10^8 bp.

The majority of human DNA exists within the nucleus of the cell. Through interactions with specific histone and other nuclear proteins, DNA is condensed into chromosomal structures. The packaging efficiency provided by these DNA:protein interactions is analogous to the packaging efficiency necessary to place approximately several hundred miles of cable wire (DNA) into the center of a basketball (nucleus). Some of these protein: DNA interactions function in structural roles, whereas others regulate expression of genes. Of the 3×10^9 bp of genomic DNA, protein-coding sequences (exons) comprise only 3–5% of the DNA. The remaining 95–97% is noncoding and includes intervening sequence (introns) and repetitive sequences; in fact, up to 50% of this noncoding DNA is some sort of repetitive sequence (Table 1) *(10)*. Although this DNA has been called "junk" sequence, evidence is building that a significant amount of this DNA has important structural and regulatory roles *(9)*. Another source of human DNA resides within the cell's mitochondria; the mitochondrial DNA consists of approx 16,500 bp that are unique to this organelle *(8)*.

DNA is considered the "blueprint" of living organisms, as it contains the information needed to create the vast array of proteins required for cellular function. The DNA sequence in an individual is the genotype; characteristics (physical, biochemical, and physiological) arising as a result of the DNA sequence interacting with the environment are referred to as the individual's phenotype. An alteration in the DNA sequence (genotype) can ultimately result in a change in the amino acid sequence of the protein encoded by that sequence; this change can render the protein nonfunctional (resulting in a specific phenotype). As our understanding of the correlation of sequence variations with altered protein functions continues to grow, nucleic acid testing will continue to play an increasing role in diagnostic laboratory medicine.

3. CATEGORIES OF GENETIC DISORDERS

Currently, there is an enormous amount of information with respect to numbers of and characteristics of various genetic

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

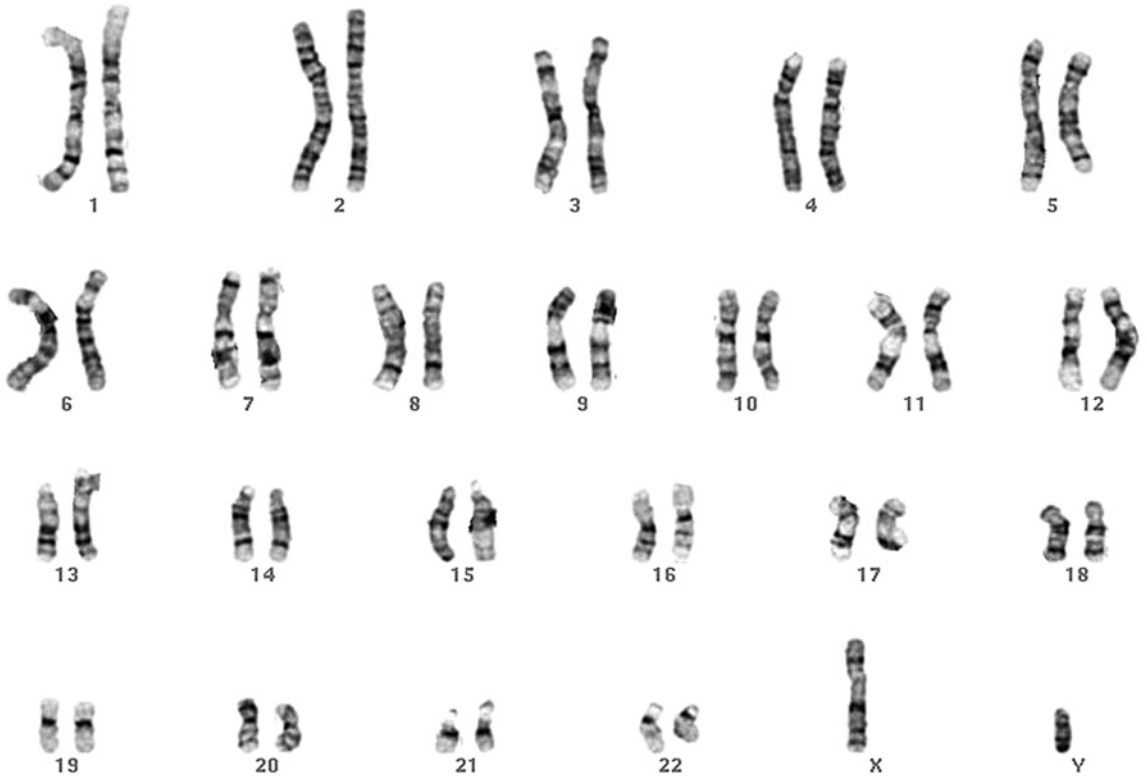


Fig. 1. A human male karyotype (G-banded) showing the 22 pairs of autosomes and the 2 sex chromosomes (XY).

Table 1
Types of Noncoding "Junk" DNA Sequence

Intron	DNA sequences that interrupt the coding (exon) sequences of a gene
Satellite	Short repetitive DNA sequences that occur at the ends (telomeres) and centers (centromeres) of a chromosome
Minisatellite	Repetitive sequences that are shorter than satellites and found throughout the genome (variable numbers of tandem repeats)
Microsatellite	Even shorter repetitive sequences: dinucleotides, trinuceotides, and tetranucleotides (short tandem repeats)
3′-Untranslated regions	DNA that is transcribed into RNA at the end of a gene but is not translated into protein; function in regulation of gene expression
Short interspersed elements (SINEs)	Well represented by the 300-bp Alu repeat that occurs approx 500,000 times within the genome
Long interspersed elements (LINEs)	Up to 700 bp in length and scattered throughout the genome

diseases and syndromes. To this list are added the growing numbers of diseases for which genetic mechanisms of disease are being identified almost daily. Genetic diseases can be categorized into three major groups: (1) chromosomal disorders, (2) monogenic or single-gene disorders, and (3) polygenic or multifactorial disorders.

Chromosomal disorders are the result of the loss, gain, or abnormal arrangement of one or more chromosomes, which results in excessive or deficient amounts of genetic material. Syndromes characterized by multiple birth defects and various forms of hematopoietic malignancy are examples of chromosomal disorders. An individual's karyotype (number and structure of chromosomes) contains 46 chromosomes; 44 of these are autosomes, designated by number from 1 to 22, and 2 are sex chromosomes designated X and Y (Fig. 1). The individual chromosomes can be distinguished from one another by size, location of centromere, and unique banding patterns after special staining methods. Chromosomal alterations usually involve large segments of DNA containing numerous genes and can be classified into four groups:

1. Aneuploidy, referring to an excess or loss of one or more chromosomes
2. Deletion, resulting from breakage and/or loss of a portion of a chromosome
3. Translocation, referring to breakage of two chromosomes with exchange of the broken parts between the chromosomes
4. Isochromosome formation, resulting from splitting of a chromosome at the centromere during mitosis, such that one arm is lost and the other duplicated to form one chromosome with identical arms

Monogenic disorders are the result of a single mutant gene and display traditional Mendelian inheritance patterns, including autosomal dominant, autosomal recessive, and X-linked. The overall population frequency of monogenic disorders is

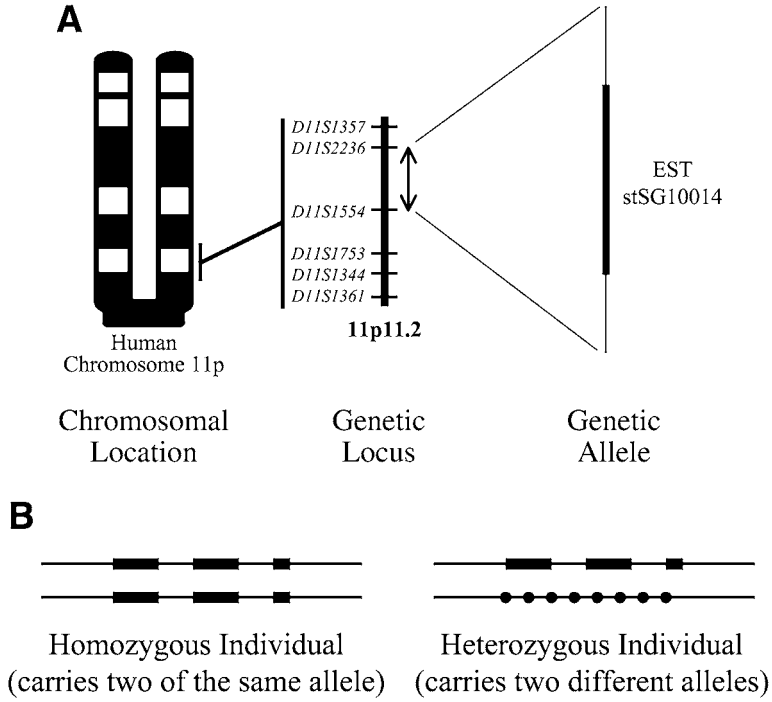


Fig. 2. (A) Schematic representation of chromosome locus and allele designation; **(B)** differentiation of a genetic locus from homozygous and heterozygous genetic alleles.

thought to be approx 10 per 1000 live births. The Human Genome Project has enabled the discovery of increasing numbers of disease-associated genes; molecular genetic testing is becoming more routine because this information is combined with advances in molecular diagnostic technologies. Biochemical lesions characteristic of monogenic disorders result from defects in a wide array of proteins, many of which are not yet characterized.

Polygenic or multifactorial disorders consist of chronic diseases of adulthood, congenital malformations, and dysmorphic syndromes. These disorders result from multiple genetic and/or epigenetic factors that do not conform to traditional Mendelian inheritance patterns. Diseases such as hypertension, ischemic heart disease, Alzheimer's disease, diabetes mellitus, and cancer develop from the interaction of numerous altered genes and environmental factors. The molecular dissection of the genetic complexity of most polygenic disorders is only in its infancy.

4. DNA SEQUENCE VARIATIONS: MUTATION AND/OR POLYMORPHISM

As noted earlier, only approx 5% of the approx 3 billion basepairs that constitute the human genome codes for proteins. Scientists have begun to understand the significance of the sequences that comprise the remaining 95%. One interesting aspect of these noncoding sequences is the amount of interindividual variation that these sequences exhibit. In some cases, these sequence variants are the result of mutation, the random alteration of DNA. When two or more sequence variants are present in a given population with a frequency of greater than 1%, then they are said to be polymorphic. At the molecular level, DNA polymorphism refers to differences in nucleotide sequences between two chromosomes at a particular genetic locus (Fig. 2A). A locus designates the position or location of a gene sequence on a chromosome; different versions of the gene at a particular locus are called alleles. Because autosomes contain two sets of chromosomes, each gene is present in two copies. Therefore, an individual can have two identical versions, or alleles, of a given gene, or can have different alleles. Homozygosity refers to having the same allele of a given gene present on both chromosomes; heterozygosity refers to having different alleles (Fig. 2B).

Genetic polymorphism is a normal occurrence, usually without grave consequences, that contributes to those traits that establish our individuality. The first types of polymorphism to be described were those that resulted in fragment length variations when DNA was digested with restriction endonucleases. These enzymes cleave DNA at specific recognition sites, producing smaller DNA fragments. Restriction fragment length polymorphisms (RFLP) occur when a restriction endonuclease recognition sequence varies between alleles at the same locus within an individual or between individuals at the same locus (Fig. 3). These variations are commonly referred to as sequence polymorphisms, because a single nucleotide within a particular DNA sequence is changed. A second type of polymorphism consists of hypervariable regions of DNA characterized by many copies of the same DNA sequence; these polymorphisms result in changes in the length of a particular sequence, as more copies are incorporated. Minisatellite sequences were the first of these types of polymorphism to be identified and are referred to as variable numbers of tandem repeats (VNTRs); VNTRs can be from 10 to 100 bp in length (Fig. 4) *(11)*. Microsatellites, or short tandem repeats (STRs), consist of dinucleotide, trinucleotide, and tetranucleotide repetitive sequences distributed throughout the genome with a frequency

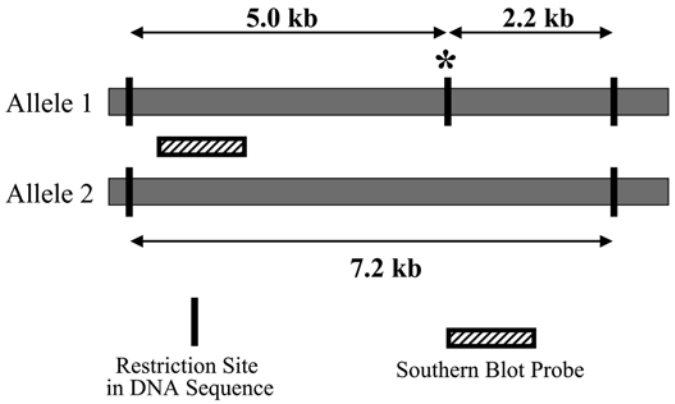


Fig. 3. Restriction fragment length polymorphism analysis for alleles 1 and 2 of a given genetic sequence. Allele 1 has restriction enzyme recognition sequences that are invariant (solid bar) and variant (*). This variant site is the site of the sequence polymorphism. When digested with the enzyme, 5.0-kb and 2.2-kb fragments are detected by gel electrophoresis and probe hybridization. In contrast, allele 2 contains only the two invariant sites. Thus, after restriction enzyme digestion, only a 7.2-kb fragment is detected.

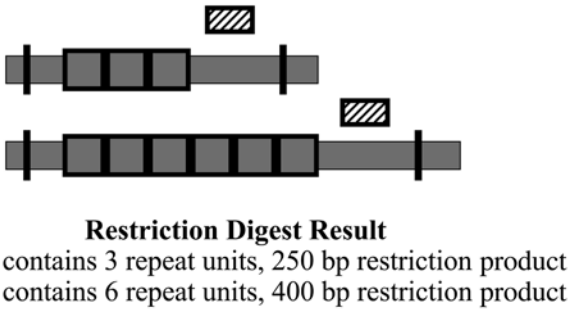


Fig. 4. Schematic diagram describing length polymorphisms of a VNTR sequence. Allele 1 contains three VNTR repeats (open boxes), whereas allele 2 only contains six repeats. These result in different size fragments, 250 bp or 400 bp, being detected by gel electrophoresis.

of approx 1 per 10 kb of sequence *(12,13).* Polymorphic DNA sequences can be used as markers for determining allelic inheritance of disease-causing genes and for identity testing.

Mutations represent permanent alterations to a DNA sequence. Common mutations, although representing sequence variants, are not usually considered polymorphisms, by virtue of the fact that they represent a rare genetic event (occurring with a frequency of less than 0.1%). Two general categories of point mutations include transitions and transversions. These mutations involve specific substitutions of the organic bases that make up a DNA sequence. The bases adenine and guanine are purine bases, whereas thymine and cytosine are pyrimidine bases. A transition mutation involves a substitution of a purine with another purine, or a pyrimidine with another pyrimidine; a transversion mutation involves the substitution of a purine for a pyrimidine or vice versa. The clinical significance of a mutation is determined by the type of mutation; its location within the gene; and the tissue involved, as not all tissues express all genes. Germline mutations are those that occur in germ cells—cells that give rise to gametes—and are, thus, heritable; these mutations will be present in every cell of the body in subsequent generations. In contrast, an acquired mutation by a cell other than a germ cell is referred to as a somatic mutation. These mutations are only passed on to daughter cells within an individual and are not heritable.

5. MENDELLIAN INHERITANCE PATTERNS

Inheritance patterns of single-gene disorders are based on traditional Mendelian laws of segregation and independent assortment. The following assumptions are made as a result of these laws: (1) An offspring inherits one autosomal chromosome from each parent and, thus, one of any given allele from each parent; (2) both alleles, regardless of inheritance, are equally expressed and heterozygotes can transmit either allele to their offspring with equal frequency; and (3) the phenotypic pattern of inheritance is dependent on whether the allele in question is located on an autosome or sex chromosome and whether the genetic disorder is dominant or recessive.

Dominant disorders are those in which a single mutant allele results in disease. Both heterozygotes and homozygotes of an autosomal dominant disorder express the disease phenotype, however, individuals homozygous for a dominant disorder are relatively rare. Recessive disorders, on the other hand, are those in which two copies of the mutant allele must be present for the disease phenotype; heterozygotes are indistinguishable from normal homozygotes. It is important to note that in recessive disorders, heterozygotes might have subtle phenotypic differences at the biochemical level that often go unnoticed. Individuals carrying one normal allele and one mutant allele are called carriers. If two individuals are carriers of alleles that cause the same recessive disorder, their offspring have a 25% chance of being affected and a 50% chance of being carriers.

The result of a mutant allele, dominant or recessive, is the result of the effect of the mutation on the role of a gene product associated with any given biological system. Disease phenotypes can be the result of the total loss or gain of protein function as a result of a single-base mutation. In some cases, a total or partial loss of protein function is observed, whereas in others, the protein functions abnormally or there is excessive normal activity. Most mutant alleles resulting in a loss of protein function exhibit recessive phenotypes. A dominant allele typically confers a gain of function, either abnormal or normal. Dominant negative phenotypes have been described, which refer to a mutant protein interfering with the normal function of the protein produced by the normal allele in heterozygotes.

The typical patterns of autosomal dominant and recessive inheritance are depicted in the following pedigrees. Autosomal dominant disorders are characterized by vertical transmission of the disease from generation to generation, equal expressivity in males and females, an affected individual having a 50% chance of offspring being affected, lack of affected children from unaffected parents, and most affected individuals having an affected parent, except in the case of a new mutation (Fig. 5). In contrast, autosomal recessive disorders are characterized by horizontal penetrance, affected homozygous individuals having unaffected heterozygous parents, and heterozygous parents having a 25% chance of offspring being affected (Fig. 6).

Several exceptions to these classic Mendelian rules of inheritance have been recognized and include recently described

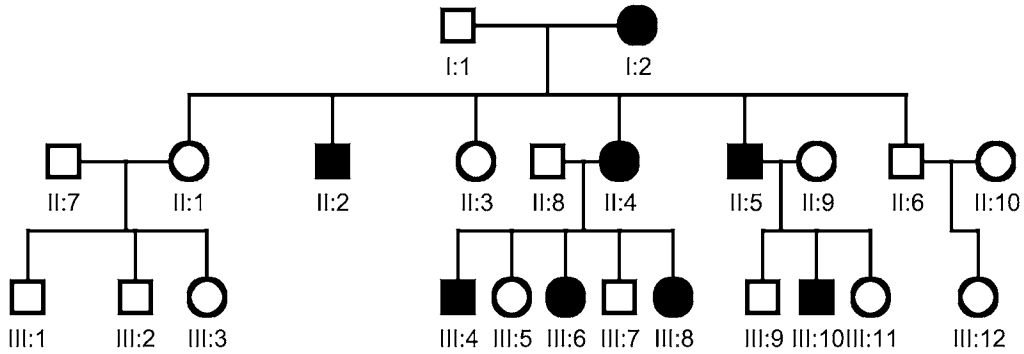


Fig. 5. Family pedigree showing autosomal dominant inheritance pattern of a genetic disease. Examples include Huntington's disease, neurofibromatosis, myotonic dystrophy, familial hypercholesterolemia, Marfan syndrome, adult polycystic kidney disease, and multiple endocrine neoplasia (MEN). Circle, female; square, male; open, unaffected; solid, affected.

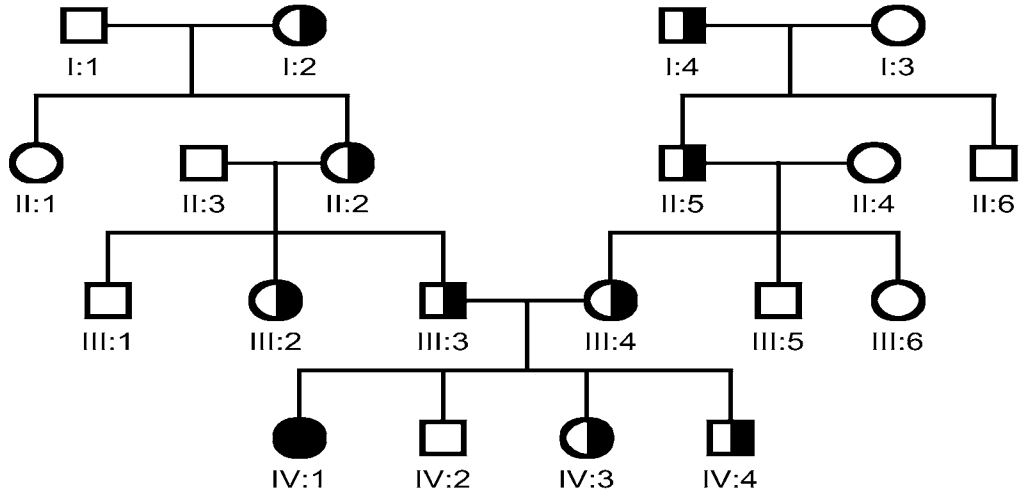


Fig. 6. Family pedigree showing autosomal recessive inheritance pattern of a genetic disease. Examples include cystic fibrosis, Tay–Sachs disease, phenylketonuria, alpha-1-antitrypsin deficiency, and sickle cell anemia. Circle, female; square, male; open, unaffected; solid, affected; half-shaded, carrier.

Table 2
Human Genes Exhibiting Differential Parental Expression

Gene	*Expressed allele*
WT-1 (Wilm's tumor suppressor)	Maternal
INS (insulin)	Paternal
IGF2 (insulinlike growth factor)	Paternal
SNRPN (small nuclear riboprotein particle)	Paternal
IGF2R (insulinlike growth factor receptor)	Maternal

Source: ref. *10.*

unstable mutations, uniparental disomy, and genetic imprinting *(5,6,13,14)*. Unstable mutations refer to the trinucleotide repeat expansions responsible for disorders such as Fragile X syndrome, myotonic dystrophy, and Huntington's disease. The identification of this type of mutation at the molecular level helped to define the phenomenon of anticipation, whereby there is an increase in the severity of a disease phentoype from one generation to the next. Premutations, a slight increase in the trinucleotide repeat number, exist without phenotypic expression. However, these premutations are prone to further expansion, which can result in a full mutation and the disease phenotype in subsequent generations. Uniparental disomy and genetic imprinting contradict assumptions that each individual inherits one copy of a single chromosome from each parent. Uniparental disomy refers to the inheritance of two copies of a chromosome from one parent and none from the other parent. This can result from differential expression of genes in one or the other parent, whereby only the expressed genes are inherited (imprinting). Parental dependency, however, can also result from the normal distribution of genetic material in male and female gametes (i.e., X vs Y chromosomes, mitochondrial genes). More commonly, parental dependent traits are the result of genetic imprinting, whereby male and female alleles are present but not equally expressed. Several human genes exhibit parental-dependent expression or imprinting, which is thought to be the result of differences in methylation patterns of specific alleles (Table 2) *(14)*.

Other patterns of inheritance include those for X-linked disorders and mitochondrial disorders. Genes responsible for X-linked disorders are located on the X chromosome. Because females have two X chromosomes and males have only one X chromosome, the severity and risk for developing these disorders is different for the two sexes. Females can be heterozygous or homozygous for a mutant X-linked gene, so that the

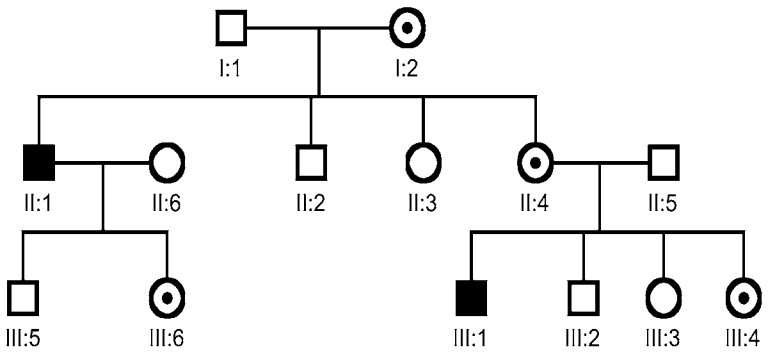


Fig. 7. Family pedigree showing X-linked recessive inheritance pattern of a genetic disease. Examples include muscular dystrophy and retinitis pigmentosa. Circle, female; square, male; open, unaffected; solid, affected; circle with dot, carrier female.

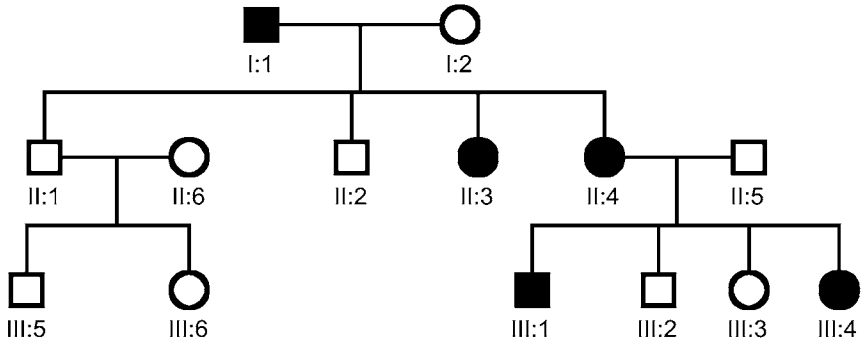


Fig. 8. Family pedigree showing X-linked dominant inheritance pattern of a genetic disease. Examples include several mental retardation syndromes. Circle, female; square, male; open, unaffected; solid, affected.

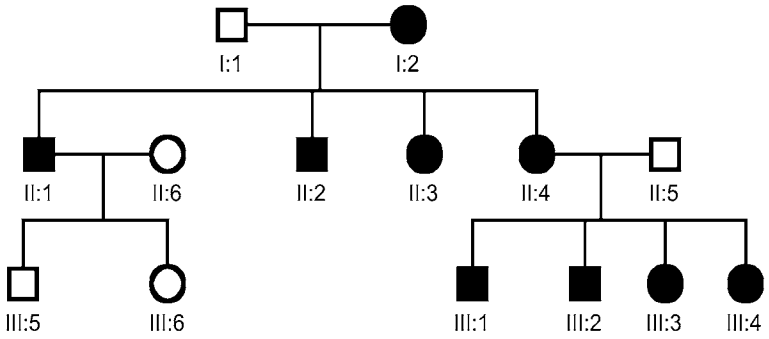


Fig. 9. Family pedigree showing mitochondrial DNA inheritance pattern of a genetic disease. Examples include Leber's hereditary optic neuropathy and Kearn–Sayre syndrome. Circle, female; square, male; open, unaffected; solid, affected.

associated trait can be either dominant or recessive. Males, on the other hand, express the mutant gene whenever they inherit it. As seen in the pedigree for X-linked disorders, there is absence of male-to-male transmission and all daughters of an affected male inherit the mutant gene. If the disorder is X-linked recessive, then affected individuals are primarily male (Fig. 7). In X-linked dominant disorders, all daughters of affected males are affected, an affected female has a 50% chance of having an affected offspring, and affected individuals have an affected parent (Fig. 8). X-inactivation, or lyonization, refers to the expression of X-linked genes in females. One X chromosome is irreversibly inactivated in females early in embryonic development and, thus, genes on only one X chromosome are expressed. Expression of a mutant allele is therefore dependent on its location on either an active or inactive X chromosome. Unlike X-linked disorders, inheritance of mitochondrial associated disorders is strictly maternal (Fig. 9). Thus, females transmit these traits to all of their offspring.

6. DIRECT VS INDIRECT MOLECULAR DIAGNOSTIC TESTS

Genetic testing, through interrogation of DNA, RNA, chromosomes, or proteins, can provide critical information for the detection of heritable disease genotypes for a number of different applications (Table 3) *(15)*. Many procedures can be used in either a direct or indirect mutation analysis. Most types

Table 3
Applications of Molecular Genetic Testing

Diagnostic testing: Testing for a gene mutation in symptomatic individuals as a diagnostic aid.

Newborn screening: Testing is used to screen populations to identify prevalent genetic mutations in asymptomatic infants. The purpose of the screening is to identify affected babies early in life to allow for appropriate intervention before irreversible damage occurs.

Presymptomatic testing: Testing for a gene mutation in asymptomatic individuals in order to predict or assess the risk of disease in the future. These applications include testing for diseases in which lifestyle changes, increased medical surveillance, or medical intervention might be beneficial if the mutation is known. Testing requires extensive pretest and posttest counseling.

Carrier screening: Testing for a gene mutation in an autosomal recessive disorder in asymptomatic individuals for the purpose of family planning and genetic counseling to determine probability of disease in children. Requires extensive pretest and posttest counseling. Carrier screening might be recommended in various situations including (1) if one or both partners have a family history of the disease, (2) if one or both partners are members of a population or ethnic group with a higher incidence of the disease, (3) if partners are seeking preconception or prenatal testing, and (4) general population screening.

Prenatal diagnosis: Testing fetal cells/tissues for mutations to determine if a fetus is affected with a disease early in the pregnancy so that termination of the pregnancy can be an option. Testing can be done at 8- to12 wk gestation by chorionic villus sampling (CVS) or at 14–16 wk by amniocentesis.

Table 4
Considerations in Genetic Test Interpretation

Penetrance: The percentage of individuals with the mutation who express the disease; not all individuals with a disease mutation will express the disease.

Heterogeneity: When a genetic disease is caused by more than one mutation. Molecular heterogeneity of a genetic disease might include mutations in a number of different genes or a large number of different mutations within the same gene. They might also result in differences in disease severity.

Expressivity: The variation in symptoms that can occur in individuals with the same genetic mutation.

Anticipation: A progressive increase in severity of a disease in future generations.

Burden of disease: The effect of a disease on the quality of life of an individual. Genetic diseases vary from those in which affected individuals can live a normal life to those that are severely disabling and ultimately fatal.

Uniparental disomy: An autosomal recessive disorder occurs when a child inherits two copies of an abnormal gene from one parent and no copies from the other parent.

Imprinting: Results in a different expression of a gene dependent on whether it was inherited from the mother or father.

of genetic testing rely on some form of in vitro amplification before proceeding with the actual analysis. Direct analysis refers to those procedures that detect the specific disease-causing mutations or foreign DNA sequence. These assays require that the mutation and/or the gene sequence of interest is known. Allele-specific oligonucleotide probes, DNA sequencing, and a wide array of polymerase chain reaction (PCR) mediated procedures are examples of direct analysis methods.

Indirect detection methods, on the other hand, are utilized when the sequence of a disease-associated gene or disease-causing mutation is not known. Polymorphic markers or gene sequences closely associated with the disease-causing gene are used to assess whether an individual has inherited the gene responsible for the disease phenotype. This type of testing is commonly referred to as linkage analysis. Linkage analysis is based on tracking the inheritance of polymorphic markers in a family with a genetic disease. If the markers and the disease-associated gene are in proximity, then the likelihood of a recombination event occurring between them is minimal. Thus, coinheritance of the markers and the disease-associated gene is likely. The advantage of linkage analysis is that the gene of interest need only be mapped to a chromosomal location. Limitations to this technology include significant labor and turnaround-times, the need to analyze samples from many family members, and the possibility of having to use numerous markers to obtain informative data. Despite technological capabilities for both direct and indirect testing, several considerations must be taken into account when interpreting these types of results (Table 4).

7. CONCLUSION

This chapter was intended to provide the reader with some of the basic concepts of molecular genetics. One should keep in mind not only the applications of this technology, which will be discussed in the following chapters, but also the "nonscience" consequences that arise as a result of genetic disease testing. Nucleic acid-based laboratory testing can be performed as a diagnostic procedure, for carrier testing, as a prenatal diagnostic test, or for presymptomatic and susceptibility testing (Table 3) *(16)*. As more of these diagnostic tests become available for genetic diseases, the laboratorian must also be aware of patterns of inheritance, risk assessment, family counseling issues, and the ethical issues associated with genetic testing. It is imperative, therefore, that the molecular genetics diagnostic laboratory function in close association with certified medical geneticists and genetic counselors. Such associations will ensure the proper dissemination and interpretation of test results, as well as maximize the benefits of genetic testing to family members.

As the identification of disease-associated mutations and genetic sequences continues to increase, so does the potential for clinical laboratorians to apply the discussed technologies to the diagnosis and monitoring of the resulting pathology in a traditional clinical laboratory fashion *(5–8)*. Health care professionals must become familiar with the advantages, disadvantages, and limitations of molecular genetic technologies when applied to the diagnosis of disease. The following chapters discuss the current applications of molecular techniques to the diagnosis of human diseases. It would be impossible, even at this early stage of molecular diagnostics, to cover all of the possibilities for molecular diagnostic testing in an up-to-date fashion. As clinical laboratorians, we must not only be aware of these rapid advances but also embrace these new technologies as the "next generation" of assays we will be performing. Rapid advances in automation promise to make

in vitro amplification, probe assays, and sequencing as routine in the clinical laboratory as the once not so popular immunoassay.

REFERENCES

1. Garrard, A. E. The incidence of alkaptonuria: a study in chemical individuality. *Lancet* **2:**1616, 1902.
2. Beadle, G. N. Biochemical genetics. *Chem. Rev.* **37:**15, 1945.
3. Watson, J. D. and Crick, F. H. C. Molecular structure of nucleic acids. *Nature* **171:**737–738, 1953.
4. Stem, R. C. The diagnosis of cystic fibrosis. N. Engl. J. Med. **336:**487–491, 1997.
5. Timchenko, L. T. and Caskey, C. T. Triplet repeat disorders: discussion of molecular mechanisms. *Cell. Mol. Life Sci.* **55:**1432–1447, 1999.
6. Jin, P. and Warren, S. T. Understanding the molecular basis of Fragile X Syndrome. *Hum. Mol. Genet.* **9:**901–908, 2000.
7. Lyon, E., and Frank, E. L. Hereditary hemochromatosis since discovery of the HFE gene. *Clin. Chem.* **47:**1147–1156, 2001.
8. Chinney, P. F. and Turnbull, D. M. Mitochondrial DNA mutations in the pathogenesis of human disease. *Mol. Med. Today,* **6:**425–432, 2000.
9. Guttmacher, A. E. and Collins, F. S. Genomic medicine: a primer. *N. Engl. J. Med.* **347:**1512–1520, 2002.
10. Nowak, R. Mining treasures from "junk" DNA. *Science* **263:** 608–610, 1994.
11. Jeffreys, A. J., Wislon, V., and Thein, S. L. Hypervariable minisatellite regions in human DNA. *Nature* **314:**67–73, 1985.
12. Lareu, M. V., Phillips, C. P., Carracedo, A., Lincoln, P. J., Syndercombe Court, D., and Thomson A. Investigation of STR locus HUMTH01 using PCR and two electrophoresis formats: UK and Galacian caucasian population surveys and usefulness in paternity investigations. *Forensic Sci. Int.* **66:**41–52, 1994.
13. Sutherland, G. R. and Richards, R. I. DNA repeats: a treasury of human variation. *N. Engl. J. Med.* **331:**191–193, 1994.
14. Barlow, D.P. Genetic imprinting in mammals. *Science* **270:**1610–1613, 1995.
15. Burke, W. Genetic testing. *N. Engl. J. Med.* **347:**1867–1875, 2002.
16. Khoury, M. J., McCabe, L. L., and McCabe, E. R. B. Population screening in the age of genomic medicine. *N. Engl. J. med.* **348:**50–58, 2003.

23 Genetic Basis of Neurologic and Neuromuscular Diseases

Narasimhan Nagan, Christopher J. Klein, D. Brian Dawson, Myra J. Wick, and Stephen N. Thibodeau

1. INTRODUCTION

The diagnosis of inherited neurological and neuromuscular disorders relies on clinical features, natural history, and the mode of inheritance. Complex factors, however, often obscure the familial tendency of these diseases and many patients go undiagnosed or are confused with noninherited disorders. Histopathologic, imaging, and electrophysiologic testing help in the identification, understanding, and characterization, but this testing is often inadequate. Molecular biology has provided for major advances in the understanding and diagnosis of these disorders. The Human Genome Project has allowed for rapid advancements in cytogenetics and other DNA testing, as both the normal and pathologic processes have been mapped, cloned, and characterized. Such work has led to specific testing and improved care of patients and their families. The types of DNA mutation implicated include trinucleotide repeat expansions, point mutations, insertions, deletions, and duplications. This chapter describes selected disorders that exemplify the types of mutation and molecular mechanism involved in neurological and neuromuscular disorders. Because molecular testing without consideration of clinical features and mode of inheritance is inadequate, a brief clinical description of these selected disorders is also provided.

2. TRINUCLEOTIDE REPEAT DISORDERS

Trinucleotide repeat expansions, initially described in 1991 *(1)*, are responsible for a number of neurological diseases, including Fragile X syndrome (FRAXA), myotonic dystrophy (DM), spinal and bulbar muscular atrophy (SBMA), Huntington's disease (1HD), Friedreich ataxia (FRDA), spinocerebellar ataxias (types SCA1, SCA2, SCA3, SCA6, SCA7, SCA8, and SCA12), and dentatorubral–pallidoluysian atrophy (DRPLA) *(2–5)*. A number of other rare forms of SCA continue to be identified in families, but commercial testing for these is currently unavailable. Most recently, SCA10 and myotonic dystrophy type 2 (DM II) have been found to result from an expansion of intronic pentanucleotide (ATTCT) and tetranucleotide (CCTG) repeats, respectively *(6,7)*. It is also worth noting that the clinical manifestation of a neurological disorder could vary with type of mutation (i.e., a trinucleotide repeat expansion vs a point mutation). This is exemplified in the case of SCA6, where point mutations in CACNL1A4 gene have resulted in familial hemiplegic migraine and episodic ataxia in contrast to the triplet repeat expansions of the same gene, which produce the spinocerebellar phenotype *(8,9)*. Because a detailed discussion of all the complexities of genetic basis of nucleotide repeat disorders is beyond the scope of our discussion, this chapter will focus mainly on the trinucleotide repeat disorders tested routinely in a clinical laboratory. Clinical and molecular features of some of the trinucleotide repeat disorders are described in the following section and are summarized in Table 1.

2.1. FRAGILE X SYNDROME Fragile X syndrome is the most frequent form of inherited mental retardation, with an incidence of about 1 in 4000 males and 1 in 6000 females. It is associated with a fragile site on Xq27.3. Fragile X syndrome exhibits X-linked dominant inheritance with reduced penetrances of 80% in males and 30% in females. Dysmorphic features associated with Fragile X syndrome include large ears, a long and narrow face, and moderately increased head circumference.

Macro-orchidism is a common finding in postpubescent affected males. Other common features include hyperactivity, attention deficit disorder, and autism.

2.2. MYOTONIC DYSTROPHY TYPE I Myotonic dystrophy type I (DM I) is the most common form of autosomal dominant adult muscular dystrophy, with a prevalence of 1 in 8000. DM is characterized by progressive muscle weakness, myotonia, cataracts, cardiac arrhythmia, and insulin-dependent diabetes. Phenotypic expression of myotonic dystrophy is variable both within and between families. Individuals who have the congenital form of myotonic dystrophy display myopathic facial appearance, respiratory distress, hypotonia, feeding difficulties, delayed motor development, and mental retardation. Increased pregnancy complications and polyhydramnios as a result of decreased fetal swallowing secondary to fetal muscle weakness are also frequently noted in cases of congenital

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

Table 1
Genetic and Molecular Characteristics of Trinucleotide Repeat Diseases

	Disorders												
	FRAXA	*DM*	*SBMA*	*HD*	*SCA1*	*SCA2*	*MJD/SCA3*	*SCA6*	*SCA7*	*SCA8*	*SCA12*	*DRPLA*	*FRDA*
Inheritance	Xlinked	AD	Xlinked	AD	AD	AD	AD	AD	AD	AD	AD	AD	AR
Chromosome location	Xq27.3	19q13.3	Xq21.3	4p16.3	6p24	12q24.1	14q32	19p33	3p12-13	13q21	5q31-33	12p12ter	9q13
Repeat sequence	CGG	CTG	CAG	CAG	CAG	CAG	CAG	CAG	CAG	CTG	CAG	CAG	GAA
Repeat location	5′UTR	3′UTR	Coding region	Coding region	Coding region	Coding region	Coding region	Coding region	Coding region	3′UTR	5′UTR	Coding region?	Intron
Gene product	FMR-1	Myotonin Protein kinase	Androgen receptor	Huntington	Ataxin-1	Ataxin-2	Ataxin-3	$_{1A}$-voltage-dependent calcium channel subunit	Ataxin-7	Unknown	PPP2R2B	Unknown	Frataxin
Normal allele	6–50	5–37	11–34	6–38	6–39	15–31	13–41	4–18	4–19	16–37	7–28	7–35	5–33
Premutation	50–200	?	—	—	—	—	—	20–23	30–36	—	—	—	34–60
Disease allele	>200	>200	36–62	27–121	40–81	36–63	61–84	21–33	37–306	110–250	66–78	49–75	66 to >1700
Transmission sex bias	Maternal	Maternal	Paternal	Paternal	Paternal	Paternal	Paternal	Unknown	Paternal	Maternal	Unknown	Paternal	Maternal

Notes: Size ranges of different alleles (normal, premutation, disease) are subject to change with additional data available.

myotonic dystrophy. Electrophysiologic testing is helpful in identifying myotonia often not recognized on clinical exam.

2.3. HUNTINGTON'S DISEASE Huntington's disease (HD) is a progressive neurodegenerative disorder affecting 1/10,000 individuals and is characterized by choreic movements, impaired cognition, and personality changes. This disorder is inherited as an autosomal dominant trait with complete penetrance and variable age of onset, with a mean age of onset at approx 40 yr. Personality changes and dementia, followed by chorea, are the presenting symptoms of the disease. Minor motor abnormalities, including clumsiness, hyperreflexia, and eye movement disturbances, are also early manifestations of HD. As the disease progresses, features of bradykinesia, rigidity, dystonia, and epilepsy could become evident.

2.4. FRIEDREICH'S ATAXIA Friedreich's ataxia (FRDA) is a slowly progressive neurodegenerative disorder affecting 1/50,000 individuals in the Caucasian population. This disorder demonstrates autosomal recessive inheritance with a variable age of onset, although most patients present with symptoms before the age of 25. It is characterized by ataxia, loss of deep-tendon reflexes, muscle weakness often associated with scoliosis, foot deformities, cardiomyopathy, and diabetes in a small fraction (10%) of affected patients.

2.5. SPINOCEREBELLAR ATAXIAS Spinocerebellar ataxias are a heterogeneous group of autosomal dominantly inherited neurologic disorders with varying levels of degeneration of the cerebellum, spinocerebellar tracts, and the brainstem neurons. Ataxia, dysarthria, and bulbar dysfunction are clinical features shared among all the SCAs. Extracerebellar features such as ocular dysfunction, extrapyramidal and pyramidal signs, peripheral neuropathy, intellectual impairment, and seizures are characteristics that are variable in presentation among the various identified ataxias. The clinically distinct Machado–Joseph disease (MJD) and spinocerebellar ataxia type 3 (SCA3) are now recognized to be allelic disorders. MJD and SCA3 are the most prevalent diseases representing 25–30% of all dominantly inherited ataxias. Both are characterized by ataxia with ophthalmoparesis and variable pyramidal–extrapyramidal findings. However, dystonia and facial fasciculations, which are present in MJD patients, are rarely observed in SCA3 patients.

2.6. X-LINKED SPINAL AND BULBAR MUSCULAR ATROPHY The X-linked recessive disorder spinal–bulbar muscular atrophy (SBMA, Kennedy's disease) is a rare adult-onset neuromuscular disorder affecting approx 1/50,000 males. The disease is characterized by the adult onset of proximal muscle weakness, atrophy, and fasciculations. Affected males often have signs of androgen insensitivity such as gynecomastia, reduced fertility, and testicular atrophy; female carriers have few or no symptoms. Diabetes is commonly seen in this disorder. The pathological findings in SBMA include degeneration of anterior horn cells and bulbar motor neurons. SBMA typically progresses slowly, and its clinical course is frequently complicated by the involvement of the bulbar muscles, but life expectancy is typically not shortened.

2.7. DENTATORUBRAL–PALLIDOLUYSIAN ATROPHY Dentatorubral–pallidoluysian atrophy (DRPLA) is a rare neurodegenerative disorder characterized by ataxia, choreoathetosis, myoclonus, epilepsy, and dementia. This disease has a variable age of onset that ranges from the first to the seventh decade.

3. GENETICS

3.1. REPEAT INSTABILITY Each of the above-described disorders is characterized by the presence of a trinucleotide repeat within the gene responsible for that disorder. For fragile X syndrome and SCA12, the trinucleotide repeat is localized to the 5′ untranslated region and for DM I and SCA8 it is present in the 3′ untranslated region. For SBMA, HD, SCA1, SCA2, SCA3/MJD, SCA6, SCA7, and DRPLA, the repeat is within the coding region. In 96% of patients with Friedreich ataxia, the repeat is located within an intron with approx 4% of the remaining affected patients, presenting as compound heterozygotes for the intronic GAA expansion on one allele and a point mutation within the Frataxin (FRDA) gene on the other allele. Common to each of these disorders is that the trinucleotide repeat is polymorphic within the normal population, with alleles inherited stably from one generation to another *(10)*. Also, in each of these disorders, expansion of the repeat beyond the normal range results in either abnormal gene function or abnormal levels of gene product and, ultimately, disease.

In the disease state, the trinucleotide repeat for each of these disorders demonstrates instability when transmitted from parents to offspring. Expansion of unstable trinucleotide repeats during transmission is most often the case, although contractions have also been documented. In general, instability of trinucleotide repeats is directly related to their size (i.e., the longer the repeat, the more likely it is to undergo expansion) *(10–12)*. This is particularly well documented for FRAXA and DM I. Stability of repeats also appears to be related to the primary sequence. In FRAXA and SCA1, for example, the trinucleotide repeat sequences are not perfect sequences, but are interrupted: interspersed AGG (instead of CGG) for Fragile X syndrome and one to three CAT repeats within the CAG repeat for SCA1 *(13,14)*. Loss of these interruption sequences appears to render the resulting repeat susceptible to greater instability with minimal expansion. For SCA1, the age of onset is determined by the number of uninterrupted CAG repeats. The presence of one or more CAT interruptions within an expanded allele have been associated with a milder phenotypic presentation as well as a later age of onset. However, when a repeat reaches a critical size threshold, the repeat becomes very unstable regardless of interruptions. For example, a repeat size of 100 or greater in FRAXA almost always leads to a full mutation (>200) in subsequent generations *(15)*. Recent studies indicate that the loss of AGG interruptions leading to an expansion of premutations to full mutations occurs as one of the final events in the expansion process that is observed in Fragile X families *(16)*.

3.2. GENOTYPE AND PHENOTYPE CORRELATION For each of the trinucleotide repeat disorders, there is a correlation between increasing repeat size and disease severity *(17,18)*. Anticipation (worsening of disease severity and decreasing of age of onset over successive generations) is well documented and correlates with increasing expansion size *(10,11,17–19)*. For example, both congenital DM patients and the most severely retarded FRAXA syndrome patients nearly always have dramatic repeat expansions, whereas those with smaller expansions typically have milder disease *(10,11,20)*. Similarly,

in individuals with HD, larger repeats are associated with prominent atrophy at the head of the caudate, a region of brain responsible for movement integration *(21)*. In case of the CAG repeat disorders, increasing repeat length correlates with earlier disease onset *(17–19,22,23)*. This correlation is strongest for SCA1, where approx 70% of the variability in age of disease onset is accounted for by repeat length *(22)*. In Friedreich ataxia, the presence of two expanded alleles as a result of the autosomal recessive nature of this disease precludes the occurrence of anticipation as seen in other dominantly inherited autosomal or X-linked trinucleotide repeat disorders. However, the age of onset and severity of disease correlate with the size of the smaller of the two repeats *(24)*. Although strong correlations exist in all triplet repeat diseases, other factors apparently also influence the severity and age onset of the diseases *(17,20,23)*. For example, in SBMA and HD, it has been reported that affected siblings with very similar repeat lengths have had onset of symptoms at very different ages *(17,23)*. Similarly, several patients with Friedreich ataxia show a later age of onset or retained reflexes despite a similar distribution of repeat sizes *(25,26)*. Thus, generally, a larger triplet repeat size correlates with more severe disease, although significant but unidentified genetic/environmental modifiers might also play a role in severity.

Fragile X syndrome, FRDA, and DM I also have repeat sizes considered to be in a "premutation" range. The premutation, which is intermediate in size between the normal repeat and "full mutation" size ranges, usually causes minimal (if any) phenotypic abnormalities. However, the premutation is unstable and often leads to further expansion and full phenotypic expression in the subsequent generation *(11)*. In addition, there is evidence for the existence of a "gray zone" in which normal and abnormal repeat sizes might overlap. The stability of repeats in this range differs between families and, thus, requires evaluation of multiple generations before repeat stability can be determined. Furthermore, stability of gray zone repeats can be affected by interruption of the trinucleotide repeat sequence, as discussed in Section 3.1.

3.3. BIAS OF PARENTAL TRANSMISSION Interestingly, parental bias has been observed with respect to trinucleotide repeat expansion in subsequent generations. For several disorders (SBMA, HD, SCA1, SCA2, SCA3/MJD, SCA7, and DRPLA), *paternal* transmission of an abnormal allele often produces expansions that are relatively large, whereas maternal transmission might result in expansions of only a few repeats *(2,19)*. Large expansions of the CAG repeat in a paternal transmission, from 43 to over 200 and from 49 to over 200, have been observed in the case of infantile onset SCA2 and SCA7, respectively *(27,28)*. These observations might reflect the potential for high repeat instability during spermatogenesis. Affected males are therefore more likely to transmit a greatly expanded repeat that could cause juvenile-onset disease. On the other hand, the untranslated CTG and CGG repeats of DM I, FRAXA, and SCA8 tend to have *maternal* bias of transmission *(29,30)*. Almost all cases of congenital myotonic dystrophy are maternally transmitted, and in FRAXA, the expansion from premutation to full mutation occurs through a oogenesis. In Friedrich ataxia, recent findings suggest that the expanded GAA alleles are likely to expand further during maternal transmission *(31)*. A uniform mechanism responsible for transmission bias of trinucleotide repeat diseases has not yet been elucidated. Prezygotic selection, at the level of DNA replication, against full mutation carrying sperm from premutation males has been proposed as leading to the observed maternal bias in transmission in case of Fragile X syndrome *(32)*.

3.4. MOLECULAR MECHANISMS The eight CAG repeat neurodegenerative disorders (SBMA, HD, SCA1, SCA2, SCA3/MJD, SCA6, SCA7, and DRPLA) are caused by modest expansions of CAG repeats that are subsequently translated into enlarged polyglutamine tracts *(2,4)*. It has been proposed that these mutations result in an *abnormal* protein that is directly responsible for the observed neuronal toxicity. This "gain of function" hypothesis is supported by the finding that patients with mutations other than (CAG) repeat expansions within the androgen receptor gene results in phenotypes (i.e., testicular feminization and androgen insensitivity syndrome but no weakness and neurodegeneration) distinct from the SBMA phenotype *(33)*. Additionally, the "gain of function" hypothesis is consistent with the dominant pattern of inheritance observed for these diseases.

Altered protein function is unlikely to be the underlying mechanism for those trinucleotide repeat diseases in which the repeat is not translated (FRAXA, DM I, and Friedreich ataxia). Rather, $(CGG)_n$ repeat expansion in the FMR-1 gene is associated with decreased mRNA and protein levels. Thus, a CGG repeat expansion greater than 200 repeats in length is associated with increased DNA methylation of an adjacent CpG island in the FMR-1 promotor region. Additionally, deacetylation of histones by histone deacetylases recruited in the vicinity of the hypermethylated FMR-1 promotor region has been shown to result in chromatin condensation and transcriptional silencing of the FMR1 gene leading to decreased protein levels *(34)*. Somatic mosaicism for FMR-1 methylation in leukocytes with fully expanded CGG repeats has been reported in high functioning Fragile X males with borderline or no mental retardation. However, because of the lack of available information on the methylation status of FMR-1 locus and FMRP expression level in the brain of these individuals, it has not been possible to extrapolate the findings of methylation mosaicism in leukocyte DNA to the severity of cognitive deficit. The appearance of a fully methylated Fragile X mutation, greater than 200 repeats, is always preceded by transmission of an unmethylated premutation, usually between 55 and 200 repeats. Premutation carriers have minimal (if any) phenotypic abnormalities. Therefore, it is likely that the decreased transcription of the FMR-1 gene in cases of fully methylated Fragile X repeat expansions is what leads to complete phenotypic expression of Fragile X syndrome. In Friedreich ataxia, the homozygous presence of expanded GAA repeats (range from 66 to greater than 1700 triplets) result in a "loss of function" because of suppression of expression of Frataxin, the FRDA gene product *(35)*. However, in contrast to Fragile X syndrome, this is not linked to an abnormal methylation of a CpG island and the mechanism(s) of transcriptional repression has not been identified *(35)*.

In DM I, it is generally believed that repeat expansion results in a reduction of steady-state myotonin protein kinase

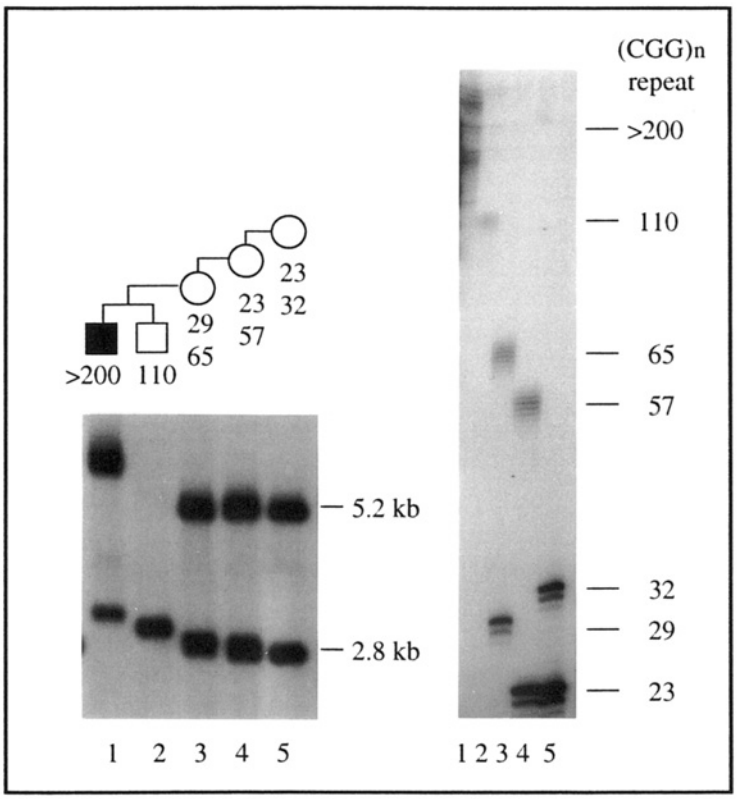


Fig. 1. Example of Southern blot and PCR analysis of DNA from individuals of a Fragile X family. **Left:** Southern blot of DNA digested with *Eco*RI and *Nru*I and hybridized with the DNA probe StB12.3; **Right:** PCR amplification of the CGGrepeat followed by denaturing gel electrophoresis. Lane numbers in the right panel correspond to those in the left. Numbers below the pedigree correspond to the CGG repeat number. The great grandmother in lane 5 has two normal alleles *(23,32)*. Her daughter (lane 4) has a normal allele of 23 and an expanded allele of 57. In subsequent transmissions, the abnormal allele of 57 expanded to 65 in one generation (lane 3) and then to 110 and >200 in the next generation (lanes 2 and 1, respectively).

(Mt-PK) mRNA and protein *(36)*. It has been proposed that the "toxic gain of function" by abnormal RNA transcripts bearing the 3´-CUG repeats, retained within the nucleus, are directly responsible for the observed neuronal toxicity through inappropriate association with other DNA/RNA or sequestration of nuclear transcription factors required for development and maturation *(37)*. An alternate model proposes chromatin restructuring caused by CTG repeats in the 3′ UTR, to result in a decreased expression of Mt-PK as well as other nearby genes *(37)*. One of the genes identified as playing a key role in the pathophysiology of myotonic dystrophy is *SIX5*, a homeobox domain gene immediately adjacent to the DMPK gene on chromosome 19q. The CTG repeat expansion in the 3′ UTR region of DMPK extends into the promotor region of *SIX5*, thereby influencing the expression of its gene product *(38)*.

3.5. MOLECULAR DIAGNOSIS The understanding of trinucleotide repeat diseases at the molecular level has had a major impact on the laboratory diagnosis of these diseases. Molecular testing has greatly improved the accuracy of the diagnosis, has allowed for presymptomatic testing of at risk family members, and has provided a means for the differential diagnosis of those diseases with overlapping clinical features. Laboratory diagnosis of the trinucleotide repeat diseases generally involves two approaches: polymerase chain reaction (PCR) and Southern blot analysis.

Primers flanking the region of DNA that contains the trinucleotide repeat are used to amplify that region by PCR. The PCR product can be analyzed by a number of techniques, including gel electrophoresis. The size of the product is determined by comparison with a standardized sizing ladder. Utilizing this approach, one can determine repeat sizes up to 200 (Fig. 1). This includes both normal and abnormal alleles of CAG repeat disorders (SCA1, SCA2, MJD/SCA3, SCA6, SCA7, SBMA, HD, and DRPLA), and normal alleles and premutations in FRAXA and Friedreich ataxia. Because the efficiency of conventional PCR is inversely correlated with the number of repeats in each allele, alleles with more than 200 repeats are more difficult to amplify and might yield no PCR product. PCR analysis, however, is simple, inexpensive, and capable of providing accurate sizing of most alleles. One of the ways to overcome this limitation in efficiency of conventional PCR is with the use of long-range PCR. This technique has been used succesfully to amplify expanded alleles (up to 1.5 kb + 3*n*, where *n* is the number of GAA repeat triplets) in Friedreich ataxia *(39)*. Many conventional PCR-based methodologies used in the detection of trinucleotide repeat expansions are transitioning into platforms that use automated fluorescence detections. These applications, when coupled with the use of chemicals such as betaine to reduce DNA secondary structure around the trinucleotide repeat expansions and *pfu* DNA

polymerase to correct for errors in replication during PCR, are expected to have a major impact in the automation of detection of trinucleotide repeat disorders in the clinical laboratory *(40)*.

In contrast to PCR, Southern blot analysis allows detection of full mutations in those diseases with large repeat expansions (FRAXA, DM I, and Friedreich ataxia) (Fig. 1). Additionally, for FRAXA, Southern blot analysis provides information concerning the methylation status of an abnormal allele. The methylation status might be of diagnostic importance when the number of repeats is near the upper end of the premutation range *(41)*. Southern blot analysis is more labor-intensive than PCR and requires larger quantities of genomic DNA. Southern blot can detect alleles in most size ranges, but in contrast to PCR analysis, it does not allow for precise sizing of trinucleotide repeats in the normal and premutation range. Full mutation alleles appear as a smear on Southern blots, because of a mosaicism in the length of the repeat in somatic cells. Radiolabeling of probes used in Southern blot analysis remains common, although nonisotopic detection methods are available as alternatives for clinical laboratories.

For diseases with small repeat expansions (SBMA, SCA1, SCA2, HD, SCA3/MJD, SCA6, SCA7, and DRPLA), PCR analysis alone is generally adequate for diagnosis. However, for those disorders with larger repeat expansions (FRAXA, DM I, and Friedreich ataxia) combined Southern blot and PCR analysis are most often used. Follow-up analysis for SCA2 and SCA7, to rule out the possibility of extreme CAG repeat expansions that have been associated with infantile or juvenile-onset ataxia *(27,28)*, is performed on samples from infants and juvenile patients identified as homozygous by PCR for a CAG repeat allele in the normal range. This is performed as a PCR–Southern blot, followed by hybridization with a $(CAG)_n$ oligonucleotide. For some of the trinucleotide repeat disorders, additional studies might be recommended. For example, the stability of an SCA1 allele with CAG repeats at the upper end of the normal range can be assessed by *Snf*II restriction endonuclease digestion, which detects the presence of CAT interruptions. In SCA1, the normal-sized (CAG) and stable gray zone alleles are nearly always interrupted by one to three CAT trinucleotide repeats, whereas the expanded or unstable gray zone SCA1 repeat is an uninterrupted sequence of CAG repeats *(14)*.

Routine cytogenetic analysis is recommended as part of a comprehensive genetic evaluation of patients referred for Fragile X syndrome testing. This testing strategy enables detection of constitutional chromosome abnormalities that might have overlapping phenotypic features of Fragile X syndrome. Finally, it should be realized that disease-causing mutations other than trinucleotide repeat expansion might occur within the aforementioned genes. Thus, the absence of trinucleotide repeat amplification does not necessarily rule out the diagnosis. This implies that for presymptomatic testing, it is important to first document the presence of a repeat amplification in an affected family member, which then verifies the underlying mechanism of disease.

4. SPINAL MUSCULAR ATROPHY

Spinal muscular atrophy (SMA) is an autosomal recessive disorder with a carrier frequency of 1/50, which affects 1/6000 to 1/10,000 live-born children. Thus, SMA is the second most common autosomal recessive lethal disease after cystic fibrosis. In this disorder, anterior horn cells degenerate, resulting in hypotonia, symmetrical muscle weakness, and wasting of voluntary muscles. The childhood spinal muscular atrophies are divided into three types (I, II, III) according to age of onset, rate of progression, and age at death. Historically, the diagnosis and classification has been made on clinical and pathological findings. Recent advances in the understanding of the genes responsible for SMA might allow confirmation of the diagnosis of SMA in symptomatic individuals and prenatal or presymptomatic diagnosis in family members.

Spinal muscular atrophy I (SMA I) or Werdnig–Hoffman disease represents 25% of all SMA cases. Onset occurs prenatally or in early infancy, with the mean onset at 1.5–3 mo. Muscle weakness and hypotonia are severe and reflexes are absent. The disease progresses rapidly with death occurring between 9 mo and 3 yr.

Spinal muscular atrophy II is an intermediate form; infants usually develop normally for 6 mo before disease onset. These infants could learn to sit alone, but do not walk unassisted. Survival into adulthood has been reported.

Spinal muscular atrophy III (Kugelberg–Welander juvenile spinal muscular atrophy) has onset between the ages of 3 and 18 yr or later and has a slower clinical course *(42)*. Atrophy and weakness of proximal muscles occur first and could be followed by distal involvement.

4.1. GENETICS All three forms of SMA are linked to 5q12-13 and might be allelic *(43)*. Four genes, the survival motor neuron gene (SMN) *(44)*, neuronal apoptosis inhibitory protein (NAIP) *(45)*, BTF2p44 *(46)*, and a putative RNA binding protein H4F5 *(47)*, are mapped to this region. Since the cloning and sequencing of SMN gene in 1996, much has been learned about the role of this gene in the genetics and pathogenesis of SMA *(48)*. A centromeric (SMN2) and a telomeric (SMN1) copy of the SMN gene is present in a 500-kb chromosomal region containing an inverted duplication *(44)*. Additionally, the genes for NAIP, BTF2p44, and H4F5 are also present as a centromeric and telomeric copy, in the same chromosomal region. The telomeric copy of SMN (SMN1) and/or NAIP, BTF2p44, and H4F5 have been shown to be deleted in affected SMA patients, although it is now well established that mutations or deletions in SMN1 are the major contributors to disease in SMA. Single-strand conformational polymorphism (SSCP) analysis of the SMN gene have shown that approx 96% of SMA type I, 94% of type II, and 82% of type III patients have homozygous deletions of exons 7 and/or 8 of the telomeric copy of the SMN gene (SMN1) *(49)*. Five percent of patients with SMA present as compound heterozygotes with a subtle intragenic mutation on one chromosome and a deletion/gene conversion on the other chromosome. In contrast to deletions, gene conversions result in an increase in centromeric SMN gene (SMN2) copy number and are associated with milder phenotypes seen in SMA III. However, there have also been reports of asymptomatic parents of affected children who have the same homozygous deletions in this region *(49)* and cases of severe forms of SMA associated with an increased SMN2 copy number. Additionally, there has been a report of one member of a sibpair having SMA I and the other having SMA III *(50)*. This

implies that there might be other modifying genes involved in the expression of SMA that influence the severity of expression of this disease. NAIP, which is located in the proximal portion of the SMA region, is reported to have deletions in 67% of type I SMA chromosomes, but only 2% of non-SMA chromosomes *(45)*. NAIP is more commonly deleted in SMA I than in types II or III *(51)*.

An emerging cellular role of SMN protein in spliceosomal snRNP biogenesis places SMA into a growing group of disorders of RNA metabolism along with Fragile X syndrome and myotonic dystrophy. This is based on the interaction of SMN with several spliceosomal Sm core proteins as well as auto oligomgerization. Exons 6 and 7 of the SMN gene contain the functionally relevant domain(s) reponsible for the auto-oligomerization of the SMN protein. Exon skipping and/or mutations within this region have been shown to result in reduced self-association and loss of protein function *(52)*. The molecular basis for the inability of SMN2 to compensate for the loss of SMN1 exon 7 has recently been understood. A C > T transition in SMN2 exon 7 disrupts an exonic splicing enhancer, a cis element that normally promotes inclusion of specific exons in pre-mRNA splicing. This results in the skipping of SMN2 exon 7 and alters the molecular structure of the SMN2 protein product *(53)*. The role of the other genes in the development of SMA is not yet clearly defined, although it has been postulated that a mutation in NAIP could lead to a failure of normally occurring inhibition of motor neuron apoptosis *(45)*.

4.2. MOLECULAR DIAGNOSIS Historically, linkage analysis was utilized for prenatal diagnosis of SMA in affected families *(54)*. Recently, identification of sequence differences has been used to distinguish SMN1 from SMN2. The sequence difference includes eight nucleotides, five of which are intronic and three of which are exonic, located within exons 6, 7, and 8 *(44,48)*. These nucleotide differences are used for the molecular diagnosis of SMA by restriction enzyme digestions. This PCR-based assay includes *Dra*I and *Dde*I digestions to distinguish between SMNI and the highly homologous SMN2. These approaches detect homozygous deletions in the SMN gene and so can be used to confirm the diagnosis in affected individuals. However, these strategies cannot be used for carrier testing or to differentiate between a true deletion and a gene conversion of SMN1 to SMN2. Use of this assay for prenatal detection of SMA is currently hampered by reports of asymptomatic individuals with homozygous deletions *(47,55)*. Testing strategies using SSCP analysis to screen for abnormalities of exons 7 and 8 in the SMN gene followed by confirmatory sequencing studies or restriction enzyme digestion to distinguish centromeric and telomeric copies of exons 7 and 8 in SMN gene have been evaluated *(55)*. However, the SSCP approach is not used because of the added disadvantage of detecting polymorphisms that resemble homozygous deletions of SMN exon 7 *(56)*.

Dosage analysis for the quantitative determination of a single-copy nondeleted SMN1 gene has recently been developed and is available in a limited number of clinical laboratories *(57)*. However, these carrier tests do not detect 4% of SMA carriers who have two copies of SMN1 on one chromosome and a null allele on the other, because they appear to be normal based on the SMN1 dosage analysis. Additionally, these assays do not identify carriers of SMA harboring subtle intragenic mutations in the SMN1 gene. Many investigators are working to fully elucidate the genetic mechanism of this disorder, which should result in improved molecular diagnosis.

5. NEUROFIBROMATOSIS 1

Neurofibromatosis 1 (NF 1) is an autosomal dominant disorder affecting approx 1/3500 individuals. It is characterized by tumors of neural crest origin. Predominant manifestations include cafe au lait spots and cutaneous or subcutaneous neurofibromas (benign tumors of peripheral nerves). Other findings include axillary freckling, Lisch nodules (iris hamartomas), scoliosis, plexiform neurofibromas, macrocephaly, short stature, seizures, and localized hypertrophy. Over half of individuals report learning problems. Patients are at increased risk for malignant tumors, including neurofibrosarcoma, astrocytoma, pheochromocytoma, embryonic rhabdomyosarcoma, and leukemia. Affected individuals could have cafe au lait spots at birth, and 97% of these individuals are symptomatic by age 20. Cutaneous neurofibromas appear in the second decade.

Neurofibromatosis 2 or central NF is a separate disorder mapped to 22q12. It is characterized by acoustic neuromas and meningiomas. Cafe au lait spots and neurofibromas are not prominent features of NF 2. This discussion will focus on NF 1.

5.1. GENETICS The gene for NF 1 is mapped to 17q11.2. The gene is extremely large, spanning 350 kb and containing at least 51 exons *(58)*. The introns range in size from 6 bp to more than 40 kb. Intron 27 contains genes for three other proteins whose role in NF 1, if any, has not been determined *(59)*. The mutation rate for the NF 1 gene is 1×10^{-4} per generation (approx 50% of the affected individuals have a new mutation).

Neurofibromin, the 2818-amino-acid protein encoded by NF 1, contains a domain that shares sequence homology with mammalian Ras p21 GTPase-activating protein (GAP) *(59)*. This domain is denoted NF1GRD for the GTPase-activity protein-related domain of the human NF type 1 protein. GAP functions to stimulate the conversion of the active GTP-bound form of Ras (p21 ras-GTP) to the inactive GDP bound form (p21 ras-GDP) *(60)*; Ras p21 proteins stimulate DNA synthesis and changes in cell morphology in response to mitogenic growth factors. Thus, NF1GRD is believed to act as a tumor suppressor gene by stimulating the conversion of p21 ras to its inactive form. Observations of increased ras activity and the loss of heterozygosity (LOH) for polymorphic DNA markers at the NF 1 locus in NF 1-associated tumors further support its role as a tumor suppressor gene *(61)*.

5.2. MOLECULAR DIAGNOSIS Screening the neurofibromin gene from affected patients has resulted in the detection of a number of mutations. The mutations include translocations, deletions, insertions, and nonsense and missense mutations *(62)*. "Hot spots" for mutation have been identified in exons 10a–10c and exon 37. Together these might account for 30% of the mutations seen in NF 1 patients *(62)*. A consortium was formed in 1993 to consolidate NF 1 mutation data *(59)*. In 2000, the consortium reported 278 mutations in screening over 500 NF 1 patients. The size of the NF 1 gene has hindered development of direct mutation analysis for use in the clinical setting.

Some laboratories have focused on the protein truncation test (PTT) for the molecular diagnosis of NF 1. The PTT assay involves generation of NF1 cDNA from cellular mRNA. This is accomplished through the use of overlapping primer sets. Coupled in vitro transcription/translation reactions generate a protein product that is then analyzed by electrophoresis. The location of the protein truncation can be estimated by comparing the migration patterns of the aberrantly migrating polypeptides with proteins of known molecular weight. Subsequent sequencing of the cDNA can be done through the use of primers designed for the region of the suspected mutation. Heim et al. *(63)* reported a 60% mutation detection rate with the NF 1 PTT assay. A significant improvement in the detection rate of NF 1 mutations (95%) was recently accomplished using a tiered approach that includes a combination of PTT, heteroduplex analysis, Southern blot, and cytogenetic analysis of patients with a possible diagnosis of NF 1 *(62)*.

Linkage analysis is available for those individuals who are part of an NF 1 pedigree. However, for the 50% of affected individuals representing sporadic cases, linkage analysis cannot be utilized. New techniques and technologies will hopefully allow for the efficient and accurate detection of mutations in affected and at-risk individuals, as well as to assist in the understanding of the function and regulation of the NF 1 gene.

6. ATAXIA TELANGIECTASIA

Ataxia telangiectasia (AT) is an autosomal recessive disorder that affects 1/40,000 to 1/100,000 individuals. It is characterized by cerebellar ataxia, oculocutaneous telangiectasias, immune defects, endocrine abnormalities, sensitivity to ionizing radiation, chromosome rearrangements, and a predisposition to malignancy. The thymus can be embryonic or hypoplastic in appearance. Defects in both humoral and cellular immunity are present and could be responsible for severe morbidity. Although prenatal growth retardation could occur, the usual onset of symptoms occurs in early childhood. Progressive ataxia is generally the first symptom, followed by choreiform movements (in approx 90% of patients) and conjunctival telangiectasias, which usually appear between 3 and 5 yr of age. AT results in a decreased life expectancy, with few patients reaching the age of 50.

The predisposition to malignancy, primarily B-cell lymphomas and chronic T-cell leukemias, is well described *(64)*. Approximately one-third of patients will develop a malignancy in their lifetime *(64)*. AT patients are also more sensitive to radiotherapy used in treatment of cancer and are at risk for early and late complications *(65)*. Interestingly, heterozygotes are also at increased risk for malignancy; increased rates of breast cancer in mothers of affected individuals have been noted *(66,67)*.

Cytogenetic evaluation provides evidence for a high degree of chromosome breakage, translocations, and inversions involving regions other than the gene locus. The translocations often involve chromosomes 7 and 14 at the sites of T-cell receptor genes and immunoglobulin heavy-chain genes *(64)*. Defective DNA repair mechanisms, including abnormal progression through cell cycle checkpoints, are thought to be responsible for the chromosome rearrangements *(64)*.

6.1. GENETICS In the late 1970s, complementation studies demonstrated the existence of at least four AT complementation groups, which suggested the possible involvement of four distinct genes. Linkage analysis, which mapped AT to 11q22-23, indicated that all four complementation groups were linked to the same locus. The gene responsible for ataxia telangiectasia, ATM, has been identified and ATM mutations have been found in all four complementation groups, suggesting a single AT gene *(68)*. ATM encodes a 12-kb transcript that has sequence homology to mammalian and yeast cell regulatory proteins *(68)*. Although these proteins appear to play a part in the cellular response to DNA damage, the specific function of the ATM gene product has not yet been elucidated.

6.2. MOLECULAR DIAGNOSIS Historically, confirmation of a clinical diagnosis of AT has been done in the cytogenetics laboratory. Bleomycin or radiation stress tests have been utilized to demonstrate the abnormal response of AT chromosomes. Karyotypes are analyzed for increased numbers of chromosome rearrangements, with particular attention paid to the chromosome 7; 14 translocations commonly seen in AT. Prenatal diagnosis in affected families has traditionally been done by linkage analysis using a fetal sample collected by amniocentesis or chorionic villi sampling *(65)*. Over 400 different mutations in the ATM gene have been identified in various populations and this figure continues to grow (http://www.vmresearch.org/atm.htm). A majority of these (70%) result in a truncated protein product and are, therefore, detected efficiently by PTT *(69,70)*. Other mutation-scanning methodologies such as SSCP and conformation-sensitive gel electrophoresis (CSGE) followed by direct sequencing have also been used with success. However, because of the heterogeneous nature of these mutations, only specific "founder" mutations in certain subpopulations such as the Amish, Moroccan Jews, Sardinians, and Mennonites are currently available for clinical testing *(71)*. Another hurdle in the identification of mutations in the ATM gene has been the difficulty in distinguishing pathological disease-causing mutations from benign polymorphisms, because of a lack of understanding of the functional domains of this protein at the cellular level.

7. HEREDITARY MOTOR AND SENSORY NEUROPATHIES

Estimated to affect 1 in 2500 individuals, Hereditary motor and sensory neuropathies (HMSNs), also termed Charcot–Marie–Tooth (CMT), are likely the most common inherited disorders of the peripheral nervous system. The classification and understanding of HMSN has evolved with each new technology, and as such, the number of distinct forms has expanded with time. Over 20 varieties with different identified loci or genes are now known *(see* Table 2). Virchow, Eichorst, and Charcot and Marie provided phenotypic accounts of the various forms of these peroneal muscular atrophies. It was on the basis of electrophysiologic, genetic, and histopathologic descriptions by Dyck and Lambert that the HMSN terminology arose *(72)*. Patients frequently have high arches, hammertoes, and prominent plantar and dorsiflexor weaknesses. Progressive worsening of handicaps is the norm, but most patients do not lose the ability to walk and often require assistive devices. In contrast,

Table 2
Hereditary Motor and Sensory Neuropathy

Disease	*Nerve conductions*	*Inheritance*	*Chromosomal location*	*Gene*
Type 1				
HMSN 1A	D	AD	17p11.2-12	PMP22
HMSN 1B	D > A	AD	1q22-23	MPZ
HMSN 1C	D	AD	16P13.1-12.3	LITAF/ SIMPLE
Type 2				
HMSN 2A	A	AD	1p35-36	?
HMSN 2B	A	AD	3q13-22	?
HMSN 2C	A	AD	12q23-24	?
HMSN 2D	A	AD	7p14	?
HMSN 2E	A > D	AD	8p21	NF-L
HMSN 2	A	AD	1q22-23	MPZ
Type 3 Dejerine-Sottas[DSS]				
HMSN 3A	D	AD	17p11.2-12	PMP22
HMSN 3B	D	AD	1q22-23	MPZ(Po)
HMSN 3C	D	AD	10q21.1-22.1	EGR2
Type 4				
HMSN 4A	D > A	AR	8q13-21.1	GDAP1
HMSN 4B1	D	AR	11q22	MTMR2
HMSN 4B2	D	AR	11p15	?
HMSN 4C	D	AR	5q32-33	?
HMSN 4D	D	AR	8q24.3	NDRG1
HMSN 4E	D	AR	10q21-22	EGR2
HMSN 4F	D	AR	19q13	PRX
Type 5				
HMSN 5	A	AD	Unknown	?
Type 6				
HMSN 6	Unclear	AR	Uknown	?
Type 7				
HMSN 7	Unclear	Unclear	Uknown	?
X-Linked HMSN				
HMSN X1	D & A	XD	Xq13.2	GJB1 (connexin 32)
HMSN X2	—	XR	Xp22.2	?
HMSN X3	—	XR	Xq26	?
HNPP				
HNPP	D	D	1711.2-12	PMP22

children who manifest with these disorders might never ambulate or lose ambulation. Breathing and bulbar weaknesses might develop in some forms of these motor and sensory neuropathies.

Increasingly molecular diagnosis is providing for classification (Table 2). Molecular genetic studies, however, without consideration of the clinical features, natural history, mode of inheritance, electrophysiologic characteristics, and neuropathological features are not as useful as desired. Specifically, genetic testing is not available in many forms and some patients, despite identical mutations, might have markedly different severity. Regardless, genetic testing is often essential, as many acquired inflammatory immune illnesses mimic these syndromes, leading to inappropriate treatment with immunosuppressants.

Beyond the scope of our discussion are those inherited neuropathies that selectively affect the sensory and autonomic fibers termed hereditary sensory and autonomic neuropathies (HSANs) *(73)*. There are five varieties described, HSAN I–V. Patients might have painless or painful injuries because of loss of sensory abilities and might have prominent autonomic dysfunction. Distinction between the forms is based on age of presentation, mode of inheritance, specific histopathologic description, and, increasingly, identification of molecular abnormalities. The mutations found are typically missense mutations. The disorders (and genes) are (1) HSAN 1 (SPTLC1), (2) HSAN II (unknown), (3) HSAN type III (IKBKAP), (4) HSAN IV (trKa), and (5) HSAN V (possibly trKa).

7.1. GENETICS The HMSN disorders are classified by whether the molecular genetic abnormality is primarily Schwann cell (myelin) or neuronal (axonal) and by the specific molecular derangements. Some disorders have unique clinical features and the classification has provided for these as separate forms. The demyelinating forms include (1) HMSN1A–C and HMSN4A–F, (2) Dejerine–Sottas syndrome (DSS also known as HMSN3A–C), (3) congenital hypomyelinating neuropathies (CHNs), and (4) hereditary neuropathy with pressure palsies (HNPPs), which is typically caused by deletions at 17p11.2, whereas HMSN1A is caused by duplications. The identified molecular defects are in structural and regulatory proteins of compact myelin, including peripheral myelin protein 22 (PMP22) *(74)* and myelin protein zero (MPZ) *(75)*. Mutations affecting noncompact myelin via gap junction protein B1 (i.e., connexin 32) (GJB1) *(76)* and the cytoskeletal-associated protein, periaxin (PRX) *(77)*, have also been identified. Proteins involved in transcription are causative and include transcription factors for late myelin genes [i.e., early growth response gene 2 {EGR2} *(78)*] and signal transduction proteins [i.e., myotubularin-related protein 2 gene {MTMR2}*(79)* and N-myc downstream-regulated gene 1 {NDRG1} *(80)*]. Finally, one gene potentially involved in regulation of apoptosis has been identified, lipopolysaccharide-induced tumor necrosis factor -α factor (LITAF) *(81)*.

Within the described genes, various micromutations could similar or dissimilar phenotypes with varied patterns of inheritance, including both dominant (PMP22, MPZ, GJB1, EGR2) and recessively (PMP22, MTMR2, NDRG1, EGR2, PRX) inherited diseases *(82)*. Of additional note is that mutations of PMP22, MPZ, and GJB1 can produce axonal diseases, implicating the critical interaction between Schwann cell elements and axonal components.

The axonal forms, HMSN 2A–E, are not as well understood as the demyelinating forms primarily because the axonal forms are not easily detected by electrophysiologic testing. The range of severity within this group of axonal disorders is broad. Abnormalities of structural proteins have been most commonly implicated, including neurofilament light chain gene (NFL) *(83)*, kinesin 1b (K1F1B) *(84)*, and gigaxonin gene (GAN1) *(85)*.

The most common mutations in the HMSN varieties are likely within PMP22. Duplication mutations of PMP22 result in HMSN 1A phenotype *(86)*, whereas deletions produce the phenotype of HNPPs *(87)*. Homologous repeat sequences (CMT1A-REP) flank the region at 17p11.2 and are thought, in part, contributory by promoting misalignment and unequal DNA recombination *(88)*. Alternate sex-linked mechanisms

exist for deletions and duplication at PMP22. *De novo* macromutations from paternal origin appear to be duplications alone. Maternal origin mutations, however, produce both duplications and deletions. The specific mutation mechanisms at PMP22 during oogenesis include unequal sister chromatid exchange and intrachromatidal loop excision *(89)*. These sex-dependent mechanisms might in part, explain the relative infrequency of HNPP compared to HMSN1A. Rare missense mutations resulting in the HMSN 1A phenotype exist *(90)*. The existence of autosomal recessive PMP22 point mutation has also been proposed, but questions about the biological significance of these basepair changes have been raised *(91)*.

PMP-22 heterozygous micromutations (typically missense mutations) can produce the severe dysmyelinating phenotype of Dejerine–Sottas syndrome (DSS) *(92,93)*. These mutations occur exclusively within the transmembrane domains and are predicted to destabilize the wild-type protein *(94)*. Factors external to PMP22 might alter clinical expression because identical mutations between and within families have been noted to produce phenotypic variability. Gene dosage has been proposed as a possible explanation for the clinical difference between HNPP (deletions) and HMSN1A (duplication) phenotypes. PMP-22 mRNA and protein levels are increased in HMSN1A and decreased in HNPP *(95,96)*. Rare exceptions are noted; for example, patients homozygous for PMP22 duplications are noted for both severe and mild phenotypes, suggesting that gene dosage alone is not sufficient to explain clinical variability *(86)*. Among the other varieties of HMSN, multiple micromutations, in connexin 32 are described, making X-linked CMT likely the second most common form of hereditary peripheral neuropathy identified to date.

7.2. MOLECULAR DIAGNOSIS One method of detecting the duplication within PMP22 is by using *Msp*I-digested DNA hybridized with the probe VAW409R3a, which maps to 17p11.2 (D17S122). This probe/enzyme combination detects three polymorphic restriction fragments and has a heterozygosity of 70%. It is interesting to note that approx 8% of CMT1A duplication patients have all three fragments *(97)*. Duplication is detected as a dosage difference between restriction fragments. A second method that offers definitive detection of CMT1A duplications is pulse-field gel electrophoresis (PFGE) of *Sac*II-digested DNA, followed by hybridization with the VAW409R3a probe. This method is likely superior, as it detects the previously described 500-kb fragment in all CMT1A duplication patients and is preferred over Southern analysis. PCR analysis utilizing simple sequence repeats in the CMT1A region is a third method for the detection of duplications. This method is informative in 80% of duplication cases, and 46% of these patients are fully informative, having three detectable alleles *(97)*. However, because it is difficult to detect duplications by PCR analysis, PCR is not the preferred method for the duplication analysis of CMT1A. Another method for the detection of duplications of the CMT1A locus is by fluorescence *in situ* hybridization (FISH) analysis with the VAW409R3a probe in combination with a control probe. Currently, the most commonly used testing strategy for detection of CMT1A duplications is Southern blot analysis, followed by PFGE in those cases in which restriction fragment length polymorphism (RFLP) analysis is uninformative *(98)*. FISH analysis, on the other hand, is recommended for the molecular diagnosis of HNPP, although an interphase FISH strategy for the detection of CMT1A duplications has recently been shown to be rapid and reliable as an alternative diagnostic approach *(87,99)*. Molecular diagnosis of CMT1A and HNPP patients with alterations other than duplication or deletion typically requires amplification and sequencing of the expressed regions by standard PCR and sequencing techniques. Such techniques are employed to identify those mutations in the remaining forms of HMSN. Commercial testing is available for connexin 32, MPZ (Po), EGR2, and NF-L, with others likely to follow.

8. DUCHENNE MUSCULAR DYSTROPHY

Duchenne muscular dystrophy (DMD) is a neuromuscular disorder that also involves gene deletion and duplication. DMD is a devastating, progressive, muscle-wasting disorder that is usually not diagnosed before the age of 3, but often results in wheelchair confinement by age 12 and death in the early twenties. Manifestations of this disease include pseudohypertrophy of muscles, joint contractures, scoliosis, respiratory compromise, cardiomyopathy, and markedly increased serum creatine kinase (CK) levels resulting from leakage of CK from diseased muscle into the bloodstream. The brain is also affected and, consequently, the IQ range of DMD patients is approx 20 points below average. However, mental abnormalities are certainly not found in all patients. The use of oral prednisone has been demonstrated to improve the strength and function of patients with DMD *(100,101)*. This observation indicates the significance of immune-modifying factors determining the clinical expression of inherited neuromuscular diseases. Aggressive symptom management might extend the survival of DMD patients into their mid-twenties. It was once believed that patients with a milder form of the disease were affected with a distinct disorder, Becker muscular dystrophy (BMD). However, molecular analysis has demonstrated that mutations in the same gene are responsible for both disorders. BMD patients could remain ambulatory through adulthood. Individuals with phenotypes intermediate to DMD and BMD have also been described *(102)*. The DMD gene has been mapped to Xq21, and although DMD is an X-linked recessive disorder, studies have demonstrated that 8% of DMD carrier females are mildly affected because of skewed X-inactivation *(103)*.

8.1. GENETICS The DMD gene was one of the first genes to be cloned by positional cloning *(104)*. The gene is extremely large, spanning approx 2300 kb, and contains 79 exons that encode the protein dystrophin. The normal protein product is not expressed in DMD patients. In skeletal muscle, dystrophin is located in the sarcolemma membrane, where it is believed to play a crucial role in linking the contractile apparatus of myocytes to the sarcolemma membrane *(105)*. Isoforms of dystrophin, produced by the use of alternative splice sites and alternative promotors, are expressed in cardiac muscle and in the brain. The finding that dystrophin is expressed in the brain offers a possible explanation for the mental subnormality often present in DMD patients. Sixty percent of DMD patients have deletions, and 6% have duplications within the DMD gene. Although the deletions are distributed throughout the

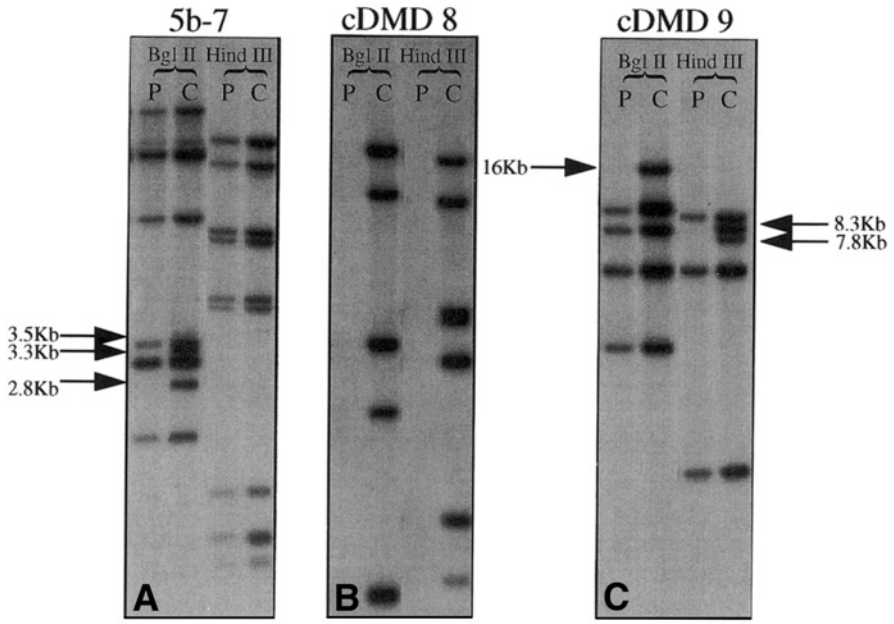


Fig. 2. Southern blot analysis of a DMD patient (P) and a normal control (C). The probe cDMD 5b–7 in combination with the *Bgl*II digest **(A)** shows absence of the 2.8-, 3.3-, and 3.5-kb bands, consistent with the deletion of exons 45–47. The probe cDMD 8 with either the *Bgl*II or the *Hin*dIII digest **(B)** shows the absence of all bands, consistent with the deletion of exons 48–52. The probe cDMD 9/*Bgl*II digest shows absence of the 16-kb band, and cDMD 9/*Hin*dIII digest shows the absence of the 7.8-kb and 8.3-kb bands (**C**); both are consistent with the deletion of exons 53 and 54. These results indicate that the deletion spans exon 45–54.

DMD gene, two deletional "hot spots" have been identified. One of these regions extends over the first 20 exons, and the second region includes exons 45–53 *(106)*. Although regions of clustered deletions suggests the presence of specific sites of breakage and recombination, such sites have not been identified. There is no correlation between the size of deletion and severity of disease. However, deletions resulting in frameshift mutations are generally associated with a severe phenotype, and most nonframe-shift mutations are associated with a BMD phenotype *(107)*. Exceptions to this rule are noted and most commonly with those patients having deletions of exons 3–7, which produces an out-of-frame mutation but leads to either DMD or Becker phenotypes *(108)*. Duplications are associated with severe phenotypes, as they often result in frameshift mutations or protein truncation. Point mutations that have been identified in DMD patients are distributed throughout exons 8–70 *(107)*.

8.2. MOLECULAR DIAGNOSIS Prior to cloning of the DMD gene and identification of dystrophin, diagnosis of DMD was based primarily on clinical features, muscle biopsy, and CK levels. Carrier testing involved examination of CK levels, which is often equivocal in carriers, and linkage analysis. Current diagnostic and carrier testing strategies include direct examination of the DMD gene and/or dystrophin analysis. Direct analysis of the DMD gene is accomplished by PCR or Southern blotting techniques. The Southern-based assay typically utilizes both *Hin*dIII and *Bgl*II restriction enzyme digestion and hybridization with the following cDNA probes: 1–2a (exons 1–9), 2b–3 (exons 10–20), 4–5a (exons 21–33), 5b–7 (exons 34–48), 8 (exons 47–52), and 9 (exons 53–59). Each exon is represented by a specific combination of bands on the autoradiogram (Fig. 2). The 3′ end of each exon has been characterized with regard to whether the exon ends with the first, second, or third nucleotide of a codon *(109)*. This information, along with Southern blot results for an individual patient, can be used to determine whether or not a particular deletion will cause a shift in the translational reading frame. Southern blot detection of carrier females and patients with duplications is based on dosage differences and requires quantitative analysis. In a small percentage of DMD patients, the junction of the deletion or duplication creates a novel restriction fragment. Carrier status determination is more easily accomplished in families where a junction fragment is present.

Polymerase chain reaction analysis typically involves multiplex reactions in which the promoter region and multiple exons are amplified (Fig. 3). The multiplexes described by Kunkel and Chamberlain *(110)* are commonly used in the clinical laboratory. PCR detects approx 98% of the deletions detected by Southern blot analysis. However, PCR-based detection of duplications and of carrier females is difficult, and PCR analysis cannot be used to distinguish between frameshift and nonframeshift deletions. It should be noted that in the prenatal setting, the speed and smaller specimen requirements of PCR might render it the method of choice for the evaluation of at risk male fetuses. For males with a suspected diagnosis of DMD, many molecular genetics laboratories begin the analysis with the PCR-based assay. If PCR fails to detect a deletion, Southern analysis can be used for further investigation. If a deletion is detected by PCR, Southern analysis with the appropriate probes can be used to confirm the extent of the deletion, as well as to distinguish between frameshift and nonframeshift mutations. The information derived from the molecular analysis can subsequently be utilized by at risk family members.

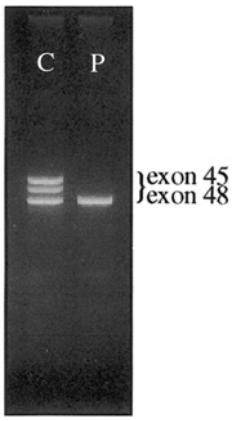


Fig. 3. PCR analysis of the DNA from a normal control (C) and the DMD patient (P) described in Fig. 2. The PCR results indicate deletion of exons 45 and 48.

Approximately 35% of DMD patients have no detectable deletions or duplications. For families of these patients, linkage analysis can often be used for prenatal evaluation and to determine the carrier status of at risk females. Linkage analysis can involve a number of approaches; the most informative are PCR-based assays in which extragenic $(CA)_n$ repeats on the 5′ and 3′ ends of the DMD gene as well as intragenic $(CA)_n$ repeats are utilized *(111)*. Although linkage analysis is highly accurate, it requires the participation of key family members, which might eliminate its utility for certain families. At-risk females who are in this situation will require alternative testing strategies for the evaluation of carrier status. This testing might include determination of CK values and dystrophin analysis. It should be noted that dystrophin analysis offers the advantage that it is not dependent on the type of mutation present in the DMD gene *(107)*. Complete sequencing of the dystrophin gene is currently being offered in a few laboratories. This will assist in the direct identification of all mutations and holds promise in the molecular diagnosis of DMD, although these techniques remain very labor-intensive. For additional information concerning DMD testing strategies, a reference by Bushby and Anderson *(111)* is recommended.

9. SUMMARY

Representative neurologic and neuromuscular genetic disorders, the disease-causing mutations, and current molecular testing strategies were discussed. The ongoing progress of the Human Genome Project will undoubtedly identify additional genes involved in neurological function. Characterization of newly identified genes, further analysis of previously identified genes, and improved technologies will allow for the design of improved molecular diagnosis and the understanding of how and possibly what important modifying factors influence the correlation of genotype to phenotype

REFERENCES

1. La Spada, A., Wilson, E., Lubahn, D., Harding, A., and Fischbeck, K. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* **352:**77–79, 1991.
2. La Spada, A. R., Paulson, H. L., and Fischbeck, K. H. Trinucleotide repeat expansion in neurological disease. *Ann. Neurol.* **36:**814–822, 1994.
3. Plassart, E. and Fontaine, B. Genes with triplet repeats: a new class of mutations causing neurological diseases. *Biomed. Pharmacother.* **48:**191–197, 1994.
4. Cummings, C. J. and Zoghbi, H. Y. Fourteen and counting: unraveling trinucleotide repeat diseases. *Hum. Mol. Genet.* **9:**909–916, 2000.
5. Durr, A. and Brice, A. Clinical and genetic aspects of spinocerebellar degeneration. *Curr. Opini. Neurol.* **13:**407–413, 2000.
6. Matsuura, T., Yamagata, T., Bugess, D., et al. Rare expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. *Nat. Genet.* **26:**191–194, 2000.
7. Liquori, C., Ricker, K., Moseley, M., et al. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science* **293:**864–867, 2001.
8. Ophoff, R. A., Terwindt, G. M., Vergouwe, M. N., et al. Familial hemiplegic migraine and episodic ataxia type-2 are caused by mutations in the Ca2+ channel gene CACNL1A4. *Cell* **87:**543–552, 1996.
9. Riess, O., Schols, L., Bottger, H., et al. SCA6 is caused by moderate CAG expansion in the alpha1A-voltage-dependent calcium channel gene. *Hum. Mol. Genet.* **6:**1289–1293, 1997.
10. Caskey, C. T., Pizzuti, A., Fu, Y. H., Fenwick, R. G., Jr., and Nelson, D. L. Triplet repeat mutations in human disease. *Science* **256:**784–789, 1992.
11. Fu, Y., Kuhl, D., Pizzuti, A., et al. Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell* **67:**1047–1058, 1991.
12. Zuhlke, C., Riess, O., Bockel, B., Lange, H., and Thies, U. Mitotic stability and meiotic variability of the (CAG)n repeat in the Huntington disease gene. *Hum. Mol. Genet.* **2:**2063–2067, 1993.
13. Snow, K., Tester, D. J., Kruckeberg, K. E., Schaid, D. J., and Thibodeau, S. N. Sequence analysis of the fragile X trinucleotide repeat: implications for the origin of the fragile X mutation. *Hum. Mol. Genet.* **3:**1543–1551, 1994.
14. Zoghbi, H. Y. and Orr, H. T. Spinocerebellar ataxia type 1. *Semin. Cell Biol.* **6:**29–35, 1995.
15. Snow, K., Doud, L. K., Hagerman, R., Pergolizzi, R. G., Erster, S. H., and Thibodeau, S. N. Analysis of a CGG sequence at the FMR-1 locus in fragile X families and in the general population. *Am. J. Hum. Genet.* **53:**1217–1228, 1993.
16. Dombrowski, C., Levesque, S., Morel, M., Rouillard, P., Morgan, K., and Rousseau, F. Premutation and intermediate-size FMR1 alleles in 10,572 males from the general population: loss of an AGG interruption is a late event in the generation of fragile X syndrome alleles. *Hum. Mol. Genet.* **11:**371–378, 2002.
17. La Spada, A., Roling, D., Harding, A., Warner, C., and Spiegel, R. Meiotic stability and genotype–phenotype correlation of the trinucleotide repeat in X-linked spinal and bulbar muscular atrophy. *Nat. Genet.* **2:**301–304, 1992.
18. Maciel, P., Gaspar, C., DeStefano, A. L., et al. Correlation between CAG repeat length and clinical features in Machado-Joseph disease. *Am. J. Hum. Genet.* **57:**54–61, 1995.
19. Nagafuchi, S., Yanagisawa, H., Sato, K., et al. Dentatorubral and pallidoluysian atrophy expansion of an unstable CAG trinucleotide on chromosome 12p. *Nat. Genet.* **6:**14–18, 1994.
20. Tsilfidis, C., MacKenzie, A. E., Mettler, G., Barcelo, J., and Korneluk, R. G. Correlation between CTG trinucleotide repeat length and frequency of severe congenital myotonic dystrophy. *Nat. Genet.* **1:**192–195, 1992.
21. Vonsattel, J., Myers, R., Stevens, T., Ferrante, R., Bird, E., and Richardson, E. Neuropathological classification of Huntington's disease. *J. Neuropathol. Exp. Neurol.* **44:**559–577, 1985.
22. Orr, H. T., Chung, M. Y., Banfi, S., et al. Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nat. Genet.* **4:**221–226, 1993.
23. Andrew, S. E., Goldberg, Y. P., Kremer, B., et al. The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. *Nat. Genet.* **4:**398–403, 1993.
24. Durr, A., Cossee, M., Agid, Y., et al. Clinical and genetic abnormalities in patients with Friedreich's ataxia. *N. Engl. J. Med.* **335:** 1169–1175, 1996

25. De Michele, G., Filla, A., Cavalcanti, F., et al. Late onset Friedreich's disease: clinical features and mapping of mutation to the FRDA locus. *J. Neurol. Neurosurg. Psychiatry* **57:**977–979, 1994.
26. Palau, F., De Michele, G., Vilchez, J. J., et al. Early-onset ataxia with cardiomyopathy and retained tendon reflexes maps to the Friedreich's ataxia locus on chromosome 9q. *Ann. Neurol.* **37:** 359–362, 1995.
27. Babovic-Vuksanovic, D., Snow, K., Patterson, M. C., and Michels, V. Spinocerebellar ataxia type 2 in an infant with extreme CAG repeat expansion. *Am. J. Med. Genet.* **79:**383–387, 1998.
28. Johansson, J., Forsgren, L., Sandgren, O., Brice, A., Holmgren, G., and Holmberg, M. Expanded CAG repeats in Swedish spinocerebellar ataxia type 7 (SCA7) patients: effect of CAG repeat length on the clinical manifestation. *Hum. Mol. Genet.* **7:**171–176, 1998.
29. Reyniers, E., Vits, L., De Boulle, K., et al. The full mutation in the FMR-1 gene of male fragile X patients is absent in their sperm. *Nat. Genet.* **4:**143–146, 1993.
30. Fu, Y. H., Pizzuti, A., Fenwick, R. G., Jr., et al. An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* **255:**1256–1258, 1992.
31. Pianese, L., Cavalcanti, F., De Michele, G., et al. The effect of parental gender on the GAA dynamic mutation in the FRDA gene. *Am. J. Hum. Genet.* **60:**460–463, 1997.
32. Moutou, C., Vincent, M. C., Biancalana, V., and Mandel, J. L. Transition from premutation to full mutation in fragile X syndrome is likely to be prezygotic. *Hum. Mol. Genet.* **6:**971–979, 1997.
33. Mhatre, A. N., Trifiro, M. A., Kaufman, M., et al. Reduced transcriptional regulatory competence of the androgen receptor in X-linked spinal and bulbar muscular atrophy. *Nat. Genet.* **5:**184–188, 1993.
34. Coffee, B., Zhang, F., Warren, S. T., and Reines, D. Acetylated histones are associated with FMR1 in normal but not fragile X-syndrome cells. *Nat. Genet.* **22:**98–101, 1999.
35. Cossee, M., Campuzano, V., Koutnikova, H., et al. Frataxin fracas. *Nat. Genet.* **15:**337–338, 1997.
36. Timchenko, L., Monckton, D. G., and Caskey, C. T. Myotonic dystrophy: an unstable CTG repeat in a protein kinase gene. *Semin. Cell Biol.* **6:**13–19, 1995.
37. Fardaei, M. and Larkin, K. Myotonic dystrophy: a multigene disorder. *Brain Res. Bull.* **56:**389–395, 2001.
38. Klesert, T. R., Otten, A. D., Bird, T. D., and Tapscott, S. J. Trinucleotide repeat expansion at the myotonic dystrophy locus reduces expression of DMAHP. *Nat. Genet.* **16:**402–406, 1997.
39. Campuzano, V., Montermini, L., Molto, et al. Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* **271:**1423–1427, 1996.
40. Hamdan, H., Tynan, J. A., Fenwick, R. A., and Leon, J. A. Automated detection of trinucleotide repeats in Fragile X Syndrome. *Mol. Diagn.* **2:**259–269, 1997.
41. Taylor, A., Safanda, J., Fall, M., et al. Molecular predictors of cognitive impairment in female carriers of Fragile X syndrome. *JAMA* **271:**507–514, 1994.
42. Baraitser, M. In *The Genetics of Neurological Disorders* 3rd ed., Baraitser, M., ed., Oxford University Press, Oxford, 1997, pp. 1770–1779.
43. Melki, J., Lefebvre, S., Burglen, L., et al. De novo and inherited deletions of the 5q13 region in spinal muscular atrophies. *Science* **264:**1474–1477, 1994.
44. Lefebvre, S., Burglen, L., Reboullet, S., et al. Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* **80:**155–165, 1995.
45. Roy, N., Mahadevan, M. S., McLean, M., et al. The gene for neuronal apoptosis inhibitory protein is partially deleted in individuals with spinal muscular atrophy. *Cell* **80:**167–178, 1995.
46. Carter, T. A., Bonnemann, C. G., Wang, C. H., et al. A multicopy transcription-repair gene, BTF2p44, maps to the SMA region and demonstrates SMA associated deletions. *Hum. Mol. Genet.* **6:**229–236, 1997.
47. Scharf, J. M., Endrizzi, M. G., Wetter, A., et al. Identification of a candidate modifying gene for spinal muscular atrophy by comparative genomics. *Nat. Genet.* **20:**83–86, 1998.
48. Burglen, L., Lefebvre, S., Clermont, O., et al. Structure and organization of the human survival motor neurone (SMN) gene. *Genomics* **32:**479–482, 1996.
49. Hahnen, E., Forkert, R., Marke, C., et al. Molecular analysis of candidate genes on chromosome 5q13 in autosomal recessive spinal muscular atrophy: evidence of homozygous deletions of the SMN gene in unaffected individuals. *Hum. Mol. Genet.* **4:**1927–1933, 1955.
50. Rudnik-Schoneborn, S., Rohrig, D., Morgan, G., Wirth, B., and Zerres, K. Autosomal recessive proximal spinal muscular atrophy in 101 sibs out of 48 families: clinical picture, influence of gender, and genetic implications. *Am. J. Med. Genet.* **51:**70–76, 1994.
51. Cobben, J. M., van der Steege, G., Grootscholten, P., de Visser, M., Scheffer, H., and Buys, C. H. Deletions of the survival motor neuron gene in unaffected siblings of patients with spinal muscular atrophy. *Am. J. Hum. Genet.* **57:**805–808, 1995.
52. Sendtner, M. Molecular mechanisms in spinal muscular atrophy: models and perspectives. *Curr. Opin. Neurol.* **14:**629–634, 2001.
53. Lorson, C., Hahnen, E., Androphy, E. and Wirth, B. A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc. Natl. Acad. Sci. USA* **96:**6307–6311, 1999.
54. Wirth, B., Rudnik-Schoneborn, S., Hahnen, E., Rohrig, D., and Zerres, K. Prenatal prediction in families with autosomal recessive proximal spinal muscular atrophy (5q11.2-q13.3): molecular genetics and clinical experience in 109 cases. *Prenat. Diagn.* **15:**407–417, 1995.
55. van der Steege, G., Grootscholten, P. M., van der Vlies, P., et al. PCR-based DNA test to confirm clinical diagnosis of autosomal recessive spinal muscular atrophy. *Lancet* **345:**985–986, 1995.
56. Wang, C. H., Xu, J., Carter, T. A., et al. Characterization of survival motor neuron (SMNT) gene deletions in asymptomatic carriers of spinal muscular atrophy. *Hum. Mol. Genet.* **5:** 359–365, 1996.
57. Ogino, S. and Wilson, R. B. Genetic testing and risk assessment for spinal muscular atrophy (SMA). *Hum. Genet.* **111:**477–500, 2002.
58. Viskochil, D., White, R., and Cawthon, R. The neurofibromatosis type 1 gene. *Ann. Rev. Neurosci.* **16:**183–205, 1993.
59. Upadhyaya, M., Shaw, D. J., and Harper, P. S. Molecular basis of neurofibromatosis type 1 (NF1): mutation analysis and polymorphisms in the NF1 gene. *Hum. Mutat.* **4:**83–101, 1994.
60. Johnson, M. R., DeClue, J. E., Felzmann, S., et al. Neurofibromin can inhibit Ras-dependent growth by a mechanism independent of its GTPase-accelerating function. *Mol. Cell. Biol.* **14:**641–645, 1994.
61. John, A. M., Ruggieri, M., Ferner, R., and Upadhyaya, M. A search for evidence of somatic mutations in the NF1 gene. *J. Med. Genet.* **37:**44–49, 2000.
62. Messiaen, L. M., Callens, T., Mortier, G., et al. Exhaustive mutation analysis of the NF1 gene allows identification of 95% of mutations and reveals a high frequency of unusual splicing defects. *Hum. Mutat.* **15:**541–555, 2000.
63. Heim, R., Kam-Morgan, L., Binnie, C., et al. Neurofibromatosis 1 truncating mutations are dispersed throughout the NF1 gene. *Am. J. Hum. Genet.* **57 (Suppl.)**, 1995.
64. Gatti, R. A. In *The Metabolic and Molecular Basis of Inherited Disease,* 8th ed., Scriver, C., Beaudet, A., Sly, W., and Valle, B., eds. McGraw-Hill, New York, 2001, pp. 709–710.
65. Gatti, R. Ataxia–telangiectasia. *Dermatol. Clin.* **13:**1–6, 1995.
66. Swift, M., Morrell, D., Massey, R. B., and Chase, C. L. Incidence of cancer in 161 families affected by ataxia-telangiectasia. *N. Engl. J. Med.* **325:**1831–1836, 1991.
67. Easton, D. Cancer risks in ataxia-telangiectasia heterozygotes. *Int. J. Radiat. Biol.* **66(6 Suppl):**177–182, 1994.
68. Savitsky, K., Bar-Shira, A., Gilad, S., et al . A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* **268:**1749–1753, 1995.
69. Teraoka, S. N., Telatar, M., Becker-Catania, S., et al. Splicing defects in the ataxia-telangiectasia gene, ATM: underlying mutations and consequences. *Am. J. Hum. Genet.* **64:**1617–1631, 1999.

70. Telatar, M., Teraoka, S., Wang, Z., et al. Ataxia–telangiectasia: identification and detection of founder-effect mutations in the ATM gene in ethnic populations. *Am. J. Hum. Genet.* **62:**86–97, 1998.
71. Gilad, S., Bar-Shira, A., Harnik, R., et al. Ataxia–telangiectasia: founder effect among north African Jews. *Hum. Mol. Genet.* **5:**2033–2037, 1996.
72. Dyck, P., Thomas, P., and Lambert, E. *Peripheral Neuropathy*. W. B. Saunders, Philadelphia, 1995.
73. Dyck, P. J. Neuronal atrophy and degeneration predominantly affecting peripheral sensory and autonomic neurons, in *Peripheral Neuropathy,* 3rd ed., Dyck, P., Thomas, P., Griffin, J., Low, P., and Poduslo, J., eds. W. B. Saunders, Philadelphia, 1993, pp. 1065–1093.
74. Naef, R. Many facets of the peripheral myelin protein PMP22 in myelination and disease. *Microsc. Res. Tech.* **41:**359–371, 1998.
75. Warner, L. E., Hilz, M. J., Appel, S. H., et al. Clinical phenotypes of different MPZ (P0) mutations may include Charcot–Marie–Tooth type 1B, Dejerine–Sottas, and congenital hypomyelination. *Neuron* **17:**451–460, 1996.
76. Nelis, E., Haites, N., and Van Broeckhoven, C. Mutations in the peripheral myelin genes and associated genes in inherited peripheral neuropathies. *Hum. Mutat.* **13:**11–28, 1999.
77. Guilbot, A., Williams, A., Ravise, N., et al. A mutation in periaxin is responsible for CMT4F, an autosomal recessive form of Charcot–Marie–Tooth disease. *Hum. Mol. Genet.* **10:**415–421, 2001.
78. Warner, L. E., Mancias, P., Butler, I. J., et al. Mutations in the early growth response 2 (EGR2) gene are associated with hereditary myelinopathies. *Nat. Genet.* **18:**382–384, 1998.
79. Bolino, A., Muglia, M., Conforti, F. L., et al. Charcot–Marie–Tooth type 4B is caused by mutations in the gene encoding myotubularin-related protein-2. *Nat. Genet.* **25:**17–19, 2000.
80. Kalaydjieva, L., Gresham, D., Gooding, R., et al. N-myc downstream-regulated gene 1 is mutated in hereditary motor and sensory neuropathy-Lom. *Am. J. Hum. Genet.* **67:**47–58, 2000.
81. Street, V. A., Bennett, C. L., Goldy, J. D., et al. Mutation of a putative protein degradation gene LITAF/SIMPLE in Charcot–Marie–Tooth disease 1C. *Neurology* **60:**22–26, 2003.
82. McKusick, V. *Mendelian Inheritance in Man,* 12th ed., Johns Hopkins University Press, Baltimore, MD, 1998.
83. Mersiyanova, I. V., Perepelov, A. V., Polyakov, A. V., et al. A new variant of Charcot–Marie–Tooth disease type 2 is probably the result of a mutation in the neurofilament-light gene. *Am. J. Hum. Genet.* **67:**37–46, 2000.
84. Zhao, C., Takita, J., Tanaka, Y., et al. Charcot–Marie–Tooth disease type 2A caused by mutation in a microtubule motor KIF1Bbeta. *Cell.* **105:**587–597, 2001.
85. Bomont, P., Cavalier, L., Blondeau, F., et al. The gene encoding gigaxonin, a new member of the cytoskeletal BTB/kelch repeat family, is mutated in giant axonal neuropathy. *Nat. Genet.* **26:**370–374, 2000.
86. Lupski, J. R., de Oca-Luna, R. M., Slaugenhaupt, S., et al. DNA duplication associated with Charcot–Marie–Tooth disease type 1A. *Cell* **66:**219–232, 1991.
87. Chance, P. F., Alderson, M. K., Leppig, K. A., et al. DNA deletion associated with hereditary neuropathy with liability to pressure palsies. *Cell* **72:**143–151, 1993.
88. Chance, P. F., Abbas, N., Lensch, M. W., et al. Two autosomal dominant neuropathies result from reciprocal DNA duplication/deletion of a region on chromosome 17. *Hum. Mol. Genet.* **3:**223–228, 1994.
89. Lopes, J., Vandenberghe, A., Tardieu, S., et al. Sex-dependent rearrangements resulting in CMT1A and HNPP. *Nat. Genet.* **17:**136–137, 1997.
90. Gabreels-Festen, A. A., Bolhuis, P. A., Hoogendijk, J. E., Valentijn, L. J., Eshuis, E. J., and Gabreels, F. J. Charcot–Marie–Tooth disease type 1A: morphological phenotype of the 17p duplication versus PMP22 point mutations. *Acta Neuropathol. (Berl.)* **90:**645–649, 1995.
91. Nelis, E., Holmberg, B., Adolfsson, R., Holmgren, G., and van Broeckhoven, C. PMP22 Thr(118)Met: recessive CMT1 mutation or polymorphism? *Nat. Genet.* **15:**13–14, 1997.
92. Ionasescu, V. V., Searby, C., and Greenberg, S. A. Dejerine–Sottas disease with sensorineural hearing loss, nystagmus, and peripheral facial nerve weakness: de novo dominant point mutation of the PMP22 gene. *J. Med. Genet.* **33:**1048–1049, 1996.
93. Ionasescu, V. V., Searby, C. C., Ionasescu, R., Chatkupt, S., Patel, N., and Koenigsberger, R. Dejerine–Sottas neuropathy in mother and son with same point mutation of PMP22 gene. *Muscle Nerve* **20:**97–99, 1997.
94. D'Urso, D., Prior, R., Greiner-Petter, R., Gabreels-Festen, A. A., and Muller, H. W. Overloaded endoplasmic reticulum-Golgi compartments, a possible pathomechanism of peripheral neuropathies caused by mutations of the peripheral myelin protein PMP22. *J. Neurosci.* **18:**731–740, 1998.
95. Yoshikawa, H., Nishimura, T., Nakatsuji, Y., et al. Elevated expression of messenger RNA for peripheral myelin protein 22 in biopsied peripheral nerves of patients with Charcot–Marie–Tooth disease type 1A. *Ann. Neurol.* **35:**445–450, 1994.
96. Schenone, A., Nobbio, L., Mandich, P., et al. Underexpression of messenger RNA for peripheral myelin protein 22 in hereditary neuropathy with liability to pressure palsies. *Neurology* **48:**445–449, 1997.
97. Wise, C. A., Garcia, C. A., Davis, S. N., et al. Molecular analyses of unrelated Charcot–Marie–Tooth (CMT) disease patients suggest a high frequency of the CMTIA duplication. *Am. J. Hum. Genet.* **53:**853–863, 1993.
98. Lupski, J. R. and Garcia, C. A. In *The Metabolic and Molecular Basis of Inherited Disease,* 8th ed., Scriver, C., Beaudet, A., Sly, W., and Valle, B., eds. McGraw-Hill, New York, 2001, pp. 5777–5778.
99. Shaffer, L. G., Kennedy, G. M., Spikes, A. S., and Lupski, J. R. Diagnosis of CMT1A duplications and HNPP deletions by interphase FISH: implications for testing in the cytogenetics laboratory. *Am. J. Med. Genet.* **69:**325–331, 1997.
100. Griggs, R. C., Moxley, T., Mendell, J. R., et al. Prednisone in Duchenne dystrophy: a randomized, controlled trial defining the time course and dose response. *Arch. Neurol.* **48:**383–388, 1991.
101. Fenichel, G. M., Florence, J. M., Pestronk, A. M., et al. Long-term benefit from prednisone therapy in Duchenne muscular dystrophy. *Neurology* **41:**1874–1877, 1991.
102. Koenig, M., Beggs, A. H., Moyer, M., et al. The molecular basis for Duchenne versus Becker muscular dystrophy: correlation of severity with type of deletion. *Am. J. Hum. Genet.* **45:**498–506, 1989.
103. Emery, A. In *Duchenne Muscular Dystrophy,* 2nd ed., Emery, A., ed. Oxford University Press, New York, 1993.
104. Monaco, A. P. and Kunkel, L. M. Cloning of the Duchenne/Becker muscular dystrophy locus. *Adv. Hum. Genet.* **17:**61–98, 1988.
105. Ervasti, J. M. and Campbell, K. P. Membrane organization of the dystrophin–glycoprotein complex. *Cell* **66:**1121–1131, 1991.
106. Koenig, M., Hoffman, E. P., Bertelson, C. J., Monaco, A. P., Feener, C., and Kunkel, L. M. Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell* **50:**509–517, 1987.
107. Worton, R., Molnar, M., Brais, B., and Karpati, G. In *the Metabolic and Molecular Basis of Inherited Disease,* 8th ed., Scriver, C., Beaudet, A., Sly, W., and Valle, B., eds. McGraw-Hill, New York, 2001, pp. 5501–5502.
108. Winnard, A., Klein, C., Coovert, D., et al. Characterization of translational frame exception patients in Duchenne/Becker muscular dystrophy. *Hum. Mol. Genet.* **2:**1347, 1993.
109. Roberts, R. G., Coffey, A. J., Bobrow, M., and Bentley, D. R. Exon structure of the human dystrophin gene. *Genomics* **16:**536–538, 1993.
110. Beggs, A. H., Koenig, M., Boyce, F. M., and Kunkel, L. M. Detection of 98% of DMD/BMD gene deletions by polymerase chain reaction. *Hum. Genet.* **86:**45–48, 1990.
111. Bushby, K. and Anderson, L., eds. *Muscular Dystrophy Methods and Protocols* Humana, Totowa, NJ, 2001.

24 Molecular Mechanisms of Endocrine Disorders

AMY POTTER AND JOHN A. PHILLIPS III

1. INTRODUCTION

The completion of the draft sequence of the human genome *(1)* promises advances in many areas of medicine, and endocrinology is no exception. Molecular diagnostic testing for RET proto-oncogene mutations has already revolutionized the management of multiple endocrine neoplasia type 2 (MEN2) and familial medullary thyroid cancer (FMTC). Other tests for genetic variation are beginning to impact the diagnosis of diabetes. Understanding the molecular mechanisms of endocrine disorders is integral to refining these and other diagnostic tests.

2. HOW MUTATIONS AFFECT THE ENDOCRINE SYSTEM

All endocrine systems consist of interacting networks of ligands, receptors, postreceptor signaling molecules, enzymes, and protein products. Each of these proteins is encoded by a gene, and these genes, in turn, are regulated by networks of transcription factors. Mutations in any of the involved genes can lead to alterations in the function of a given system and, thus, to disease. Selected endocrine disorders that illustrate particular molecular mechanisms of disease will be discussed here; other disorders are discussed in more detail later in the chapter.

A variety of mutations have been reported in the genes encoding hormones themselves. Deletions of such genes will result in complete absence of hormone, whereas less drastic mutations will result in a nonfunctional protein product. Mutations have also been reported that alter processing and packaging of the hormone. One example is the AVP gene (OMIM 192340), which encodes both the hormone arginine vasopressin (AVP) and its packaging protein neurophysin II. Mutations within the neurophysin II domain impair packaging and secretion of AVP, leading to central diabetes insipidus *(2)*.

Mutations in cell surface receptors are another important genetic cause of endocrine disorders. The severity and manifestations of an illness can vary depending on whether mutations result in no receptor function (null mutations) or only in decreased functioning. Receptor mutations can lead to decreased ligand binding or impaired signaling. For example, null mutations of the insulin receptor lead to leprechaunism (also called Donohue syndrome, OMIM 246200); these patients have nearly absent insulin function and die in infancy *(3–6)*. Less severe mutations in the insulin receptor cause Rabson–Mendenhall syndrome (OMIM 262190), which is associated with survival into late childhood and early adolescence *(7)*.

Two groups of hormones, the steroid hormones and the thyroid hormones, interact with nuclear receptors. In order to function properly, the receptor must be able to bind both its ligand and its appropriate DNA-binding site, as well as to interact with other cofactors. For example, androgen insensitivity syndrome is caused by mutations in the androgen receptor (OMIM 313700); the majority of such mutations are in the ligand-binding domain, but up to 20 % occur in the DNA-binding domain *(8,9)*.

Mutations in the proteins of post-receptor signaling systems can also cause endocrine disease. The classic example is the McCune–Albright syndrome (OMIM 174800). This syndrome is caused by postzygotic somatic mutations in the α-subunit of the adenylate cyclase stimulatory G-protein ($G_s\alpha$, encoded by GNAS1, OMIM 139320). $G_s\alpha$ is a key regulator of the hormone-responsive adenylate cyclase system and is involved in intracellular signaling by many endocrine receptors, including those for adrenocorticotropic hormone (ACTH), thyroid-stimulating hormone (TSH), parathyroid hormone (PTH), leutinizing hormone (LH), and follicle-stimulating hormone (FSH). Activating mutations lead to constitutive activity of these receptors. The classic triad of McCune–Albright syndrome consists of polyostotic fibrous dysplasia, precocious puberty, and café au lait spots, but hyperfunction of multiple endocrine systems can occur including hyperthyroidism, hypercortisolism, growth hormone excess, and hyperprolactinemia. The timing of the somatic mutation causes variation in the type and extent of manifestations. The majority of patients have a point mutation in exon 8 of the GNAS1 gene (Arg201His or Arg201Cys), although individuals with other mutations have been reported. Molecular analysis is best done on affected tissue, as unaffected tissue might not show the mutation *(10–12)*.

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

Many endocrine disorders result from defects in the enzymes that synthesize the respective hormones. Deficiencies in any of the several enzymes involved in adrenal steroid hormone biosynthesis can result in congenital adrenal hyperplasia, whereas defects in thyroid hormone biosynthesis can result in congenital hypothyroidism.

Finally, descriptions of how defects in various transcription factors can cause clinical disease are provided in the following sections on the growth hormone pathway and maturity-onset diabetes of youth (MODY).

3. MOLECULAR GENETICS IN ENDOCRINOLOGY

3.1. THE GROWTH HORMONE PATHWAY The growth hormone (GH) pathway is comprised of a series of interdependent genes whose products are required for normal growth (see Fig. 1). The GH pathway genes include ligands (GH and insulin-like growth factor 1 or IGF-1), transcription factors (PIT1 and PROP1), agonists (GH-releasing hormone or GHRH), antagonists (somatostatin), and receptors such as the GHRH receptor (GHRHR) and the GH receptor (GHR). These genes are expressed in different organs and tissues, including the hypothalamus, pituitary, liver, and bone (see Fig. 1). The effective and regulated expression of the GH pathway is essential for growth in stature as well as for homeostasis of carbohydrate, protein, and fat metabolism. Diseases caused by known gene defects in the GH pathway include combined pituitary hormone deficiency (CPHD) caused by defects in transcription factors such as HESX1, LHX3, PROP1 and PIT1, isolated growth hormone deficiency (IGHD) caused by defects in the GH gene (GH1) and in GHRHR, and GH-resistance syndromes caused by defects in GHR and IGF1.

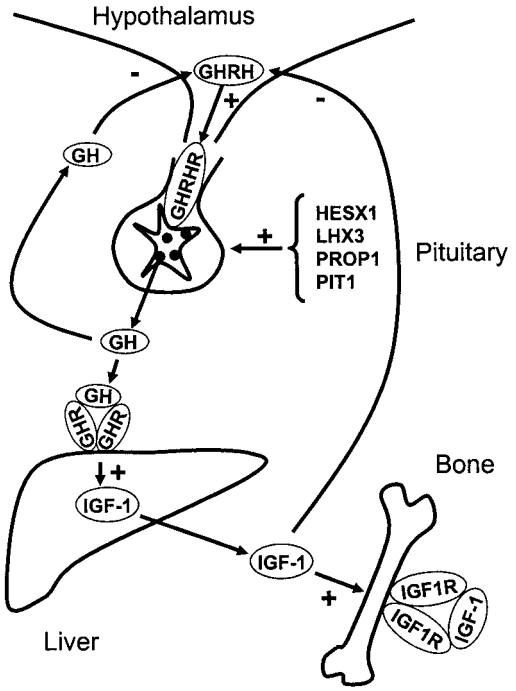


Fig. 1. The growth hormone pathway. For details, see text. GH: growth hormone; HESX1: homeobox gene expressed in ES cells; LHX3: lim homeobox gene 3; PROP1: prophet of PIT1; PIT1: pituitary-specific transcription factor 1; GHRH: growth hormone-releasing hormone; GHRHR: growth hormone-releasing hormone receptor; GHR: growth hormone receptor; IGF-1: insulin-like growth factor 1; IGF-1R: insulin-like growth factor 1 receptor.

3.2. ISOLATED GROWTH HORMONE DEFICIENCY Estimates of the frequency of GH deficiency (GHD) range from 1 in 4000 to 1 in 10,000 in various studies. Causes of GHD include central nervous system insults or defects such as cerebral edema, chromosome anomalies, histiocytosis, infections, radiation, septo-optic dysplasia, trauma, or tumors that affect the hypothalamus or pituitary. Although most individuals with GHD have no family history of GHD, estimates of the proportion of cases that have an affected parent, sibling, or child range from 3% to 30% in different studies. This familial clustering suggests that a significant proportion of GHD cases might have a genetic basis and that having an affected relative conveys substantially increased relative risk for all first degree relatives *(13–14)*. Our current understanding of the genetics of familial defects in the GH pathway will be discussed in the following subsections.

3.3. CLINICAL FEATURES An intact GH pathway is needed throughout childhood to maintain normal growth. Concomitant or combined deficiencies of other pituitary hormones (LH, FSH, TSH, and/or ACTH), in addition to GH, is called combined pituitary hormone deficiency (CPHD) or panhypopituitary dwarfism. The combination of GH and these additional hormone deficiencies often causes more severe retardation of growth, and skeletal maturation and spontaneous puberty might not occur *(13,14)*.

3.4. DIAGNOSIS Short stature, delayed growth, and delayed skeletal maturation occur with GH pathway defects. Because these signs can also be associated with systemic illnesses, individuals suspected of having GHD should be evaluated for systemic diseases before having complicated tests to detect GHD. GH stimulation testing is necessary to confirm the diagnosis of GHD, and peak GH levels of less than 7–10 ng/mL are considered abnormal. Testing for concomitant deficiencies of LH, FSH, TSH, and/or ACTH should be done when GHD is diagnosed to detect CPHD and provide a complete diagnosis and enable planning of optimal treatment *(13,14)*.

3.5. TYPES OF FAMILIAL IGHD There are at least six different Mendelian disorders. These include four autosomal recessive disorders (IGHD IA and IB, bioinactive GH, and GHRHR defects; OMIM 262400, 139250, 262650, and 139191, respectively). In addition, there are autosomal dominant (IGHD II, OMIM 173100) and X-linked forms of IGHD (IGHD III, OMIM 307200) (Table 1).

3.5.1. Autosomal Recessive Isolated Growth Hormone Deficiency The most severe form of IGHD, called IGHD IA (OMIM 262400 and139250) has an autosomal recessive mode of inheritance. Affected neonates occasionally have mildly decreased birth lengths and hypoglycemia in infancy. All develop severe dwarfism by 6 mo of age. Although replacement therapy with recombinant human growth hormone (rhGH) gives a good initial growth response in individuals with IGHD IA, this response is often temporary because of formation of anti-GH antibodies, which, in sufficient titer, cause GH resistance and an arrest of the response to rhGH replacement.

Table 1
Selected Genetic Defects Causing Isolated Growth Hormone Deficiency

Type of IGHD	*Location*	*Nucleotide change*	*Effect*
IGHD IA		Deletion	6.5-kb deletion of GH1 gene
		Deletion	7.0-kb deletion of GH1 gene
		Deletion	7.6-kb deletion of GH1 gene
		Deletion	45-kb deletion of GH1 gene
	Exon 2	5536delC	Frameshift after 17th amino acid of signal peptide
	Exon 2	5543G > A	Stop codon after 19th amino acid of signal peptide
IGHD IB	IVS 4	+1G > C	Donor splice site mutation; retains exons 1–3, part 4
		+1G > T	Donor splice site mutation; retains exons 1–3, part 4
		+5G > C	Donor splice site mutation; retains exons 1–3, part 4
IGHD II	Exon 3	+5G > A	Exon splice enhancer mutation; exon 3 skip
	IVS 3	+1G > A	Donor splice site mutation; exon 3 skip
		+1G > C	Donor splice site mutation; exon 3 skip
		+2T > C	Donor splice site mutation; exon 3 skip
		+5G > A	Donor splice site mutation; exon 3 skip
		+5G > C	Donor splice site mutation; exon 3 skip
		+6T > C	Donor splice site mutation; exon 3 skip
	IVS 3	+28G > A	Intron splice enhancer mutation; exon 3 skip
		Del 28-45	Intron splice enhancer mutation; exon 3 skip
	Exon 5	6664G>A	Arg183His; altered secretory granules

Isolated growth hormone defeciency IA is usually caused by deletions of the GH1 gene. At a molecular level, these DNA deletions are usually 6.7, 7.0, or 7.6 kb in length, with approx 75% being 6.7 kb *(13,14)*. Several additional cases with 45 kb deletions have also been reported in unrelated families from Turkey, Italy, and Asia (see Table 1) *(14)*. DNA sequence analysis of the fusion fragments associated with the recurring 6.7 to 7.6 kb deletions has shown that they arose from homologous recombination between repeated sequences that flank the GH1 gene *(15)*. Multiple studies indicate that about 15% of individuals with severe IGHD (-4.5 SD in height or lower) have GH1 gene deletions. In 10 GH gene deletion subjects identified in studies of 78 subjects with severe IGHD, 3 out of 10 (30%) developed anti-GH antibodies following treatment with rhGH *(16)*.

Sequencing studies have detected point mutations causing IGHD-IA *(17–19)*. In a consanguineous Turkish family, a G>A transition in the 20th codon of the GH1 signal peptide was found that converts a TGG (Trp) to a TAG (Ter) codon. In another patient, Duquesnoy et al. found a deletion that was inherited from one parent and a frameshift in the signal peptide of the GH1 gene that was inherited from the other (Table 1) *(17)*. Because frameshift and nonsense mutations, as well as gene deletions, have been found to cause the IGHD IA phenotype, this disorder is best described as complete GHD because of heterogeneous GH1 gene defects, rather than gene deletions alone (Table 1) *(13,14)*.

A second autosomal recessive type of IGHD (IGHD IB) is characterized by the production of deficient but detectable amounts of GH after provocative stimuli. This contrasts with the absence of GH secretion that occurs in IGHD IA. In contrast to IGHD IA subjects, those with IGHD IB have no detectable GH1 gene deletions *(20)*. The clinical criteria used to classify the IGHD IB phenotype include an accelerated growth response to rhGH *(20)*. This observation and lack of anti-GH antibodies in those studied suggests that sufficient endogenous GH secretion occurs to prevent the anti-GH antibody production that characterizes IGHD IA. Interestingly, some subjects with IGHD IB have been shown to have intact secretory granules in their somatotropes and to exhibit normal GH responses to GHRH infusions *(21,22)*. These observations suggest that the GH1 allele(s) are capable of some expression and that defects in GHRH or GHRHR synthesis or secretion might underlie this disorder. Using DNA sequence analysis of polymerase chain reaction (PCR) amplification products, Cogan et al. detected a point mutation causing IGHD IB *(19)*. In a Saudi Arabian family, a G>C transversion was found that alters the first base of the 5′ splice site (5′SS) of intron 4 (IVS4). This and other substitutions of the 5′SS perturb GH1 mRNA splicing and cause IGHD IB (Table 1). Abdul-Latif et al. identified a G>C transversion of the fifth base of the 5′ SS of IVS4 in a consanguineous family with IGHD IB (Table 1) *(23)*. Reverse transcription (RT)–PCR transcripts from lymphoblasts of an affected patient demonstrated that the IVS4 +5G>C had the same effect on splicing as the IVS4 +1G>C. A fourth IGHD IB patient with severe growth retardation was found to have two different GH gene defects by Igarashi et al. One allele had a 6.7-kb GH gene deletion and the other had a 2-bp deletion in exon 3 (Table 1) *(24)*. The 2-bp deletion causes a frameshift within exon 3, generating a premature stop codon within exon 4. The patient responded well to rhGH replacement therapy and did not produce anti-GH antibodies.

3.5.2. Autosomal Dominant Isolated Growth Hormone Deficiency A third type of IGHD (IGHD II) has an autosomal dominant mode of inheritance. Patients diagnosed with IGHD II have a single affected parent and respond well to rhGH. Patients differ in the severity of GH deficiency and their propensity for the development of hypoglycemia *(25)*. Most of the IGHD II GH1 gene defects reported to date are mutations that alter splicing of GH mRNA and cause skipping of exon 3 (Table 1).

Table 2
Known Genes Causing MODY

	Gene	*OMIM*	*Treatment*	*Complications*	*Additional features*
MODY1	HNF4α	600281	Yes	Yes	None
MODY2	GCK	138079	Often no	No	Permanent insulin-requiring neonatal diabetes in homozygotes
MODY3	HNF1α	142410	Yes	Yes	Low renal threshold for glucose resulting in glycosuria
MODY4	IPF1	600733	Yes	Probably	Pancreatic agenesis and permanent insulin-requiring neonatal diabetes in homozygotes
MODY5	HNF1β	189907	Yes	Probably	Renal cysts, progressive non-diabetic renal dysfunction leading to renal failure, genital abnormalities in females
MODY6	NeuroD1 (BETA2)	601724	Yes	Probably	None

The first GH1 gene defect was identified in an IGHD II kindred that showed segregation with a marker tightly linked to the GH1 gene (Table 1) *(19)*. DNA sequence analysis of PCR amplification products of four affected relatives showed a change in the sixth base of the IVS3 5′ SS (Table 1). This mutation causes skipping or deletion of exon 3, which codes for amino acids 32–71, including one of the four cysteines in GH necessary for intramolecular disulfide bonding. The resulting truncated GH protein product retains the signal peptide and exon 4–5 sequences thought to be important for transport to the secretory granules. The presence of the mutant and normal GH protein products within the secretory granules results in derangements and apoptosis of somatotropes that inhibit expression of the normal GH molecule. Recently, the mechanism by which this mutant GH protein interferes with the normal GH protein has been studied in mice transgenic for another mutation (IVS3 + 1 G > A transition) that causes IGHD II. McGuinness et al. found that expression of the mutant protein disrupts secretory vesicles and causes widespread pituitary damage *(26)*.

Two additional IGHD II mutations alter the first base of the IVS3 5′ SS. One is a G > C transversion (IVS3 + 1G > C) and the other is a recurring G > A transition (IVS3 + 1G > A) (Table 1) *(27,28)*. The latter was found in three nonrelated families and is thought to be caused by the high mutation frequency at CpG dinucleotides that result in C to T and G to A transitions (discussed further under GHR point mutations). Both IGHD II mutations were proven to cause exon 3 skipping in lymphoblastoid cells and transfection studies, respectively. Other IGHD II mutations in IVS3 that also cause skipping of exon 3 but do not occur within the branch consensus, 5′ SS or 3′ SS have been identified *(29)*. The first deletes 18 bp (IVS3 del+ 28–45) and the second is a G > A transition (IVS3 + 28G > A) (Table 1). RT-PCR amplification of GH1 gene transcripts from transient expression studies, using IGHD II mutant constructs in mammalian cells, yielded DNAs showing dramatic increases in exon 3 skipping relative to normal controls. The shift toward exon 3 skipping is caused by the disruption of an intronic XGGG repeat, which is an intron splicing enhancer (ISE) that regulates the pattern of alternative splicing of GH mRNA.

Recently, an additional GH1 mutation was reported that causes IGHD II by its effects on an exon splice enhancer (ESE). Moseley et al. reported an A > G transition of the fifth base of exon 3 (E3 + 5 A > G) in affected individuals from an IGHD II family. This mutation disrupts a $(GAA)_n$ ESE motif immediately following the weak IVS2 3′ SS. To determine the effect of ESE mutations on GH mRNA processing, GH3 cells were transfected with expression constructs containing either the normal ESE, +5 A > G, or other ESE mutations, and cDNAs derived from the resulting GH mRNAs were sequenced. All ESE mutations studied reduced activation of the IVS2 3′ SS and caused either partial E3 skipping, because of activation of an E3 +45 cryptic 3′ SS, or complete E3 skipping. Partial or complete E3 skipping led to loss of the codons for amino acids 32–46 or 32–71, respectively, of the mature GH protein. The authors conclude that the E3 +5 A > G mutation causes IGHD II because it perturbs an ESE required for GH splicing *(30)*.

One IGHD II mutation has been reported that does not occur within IVS3 or affect splicing. This mutation is a G > A transition that results in an Arg > His substitution at residue 183 (Arg183His) of the mature GH protein (Table 1) *(31)*. This substitution is believed to alter the intracellular processing of the normal GH molecules by binding zinc and thereby deranging the zinc-associated presecretory packaging of GH.

3.5.3. X-Linked Isolated Growth Hormone Deficiency

A third form of IGHD called IGHD III (OMIM 307200) has an X-linked mode of inheritance and distinct clinical findings in different families. In some families, all cases have agammaglobulinemia associated with their IGHD, whereas in other families, all cases have only IGHD. This suggests that contiguous gene defects on the long arm of the X chromosome might cause some IGHD III cases. Duriez et al. reported that X-linked agammaglobulinemia and IGHD is caused by mutation in the Bruton's tyrosine kinase or BTK gene *(32)*.

Laumonnier et al. studied the SOX3 gene in families with X-linked mental retardation and GH deficiency where the causative gene had been mapped to Xq26-q27. They showed that the SOX3 gene maps to Xq26.3 and was involved in a large family in which affected individuals had mental retardation and IGHD (OMIM 300123 and 313430). The mutation was an in-frame duplication of 33 bp encoding 11 alanines in a polyalanine tract of the SOX3 gene. The expression pattern during neural and pituitary development suggested that dysfunction of the SOX3 gene caused by this polyalanine expansion might disturb transcription pathways and the regulation of genes involved in pituitary development *(33)*.

3.5.4. Biodefective Growth Hormone (Autosomal Recessive) Some children with short stature comparable to that seen in GH deficiency have low levels of IGF-I but normal levels of GH assayed by radioimmunoassay (RIA) *(34,35)*. In such cases, the administration of rhGH is reported to produce an increase in IGF-I levels and a growth response. In some, the concentration of GH as measured by RIA greatly exceeds the concentration measured by radioreceptor assay. These results suggest that the primary defect could be production of an abnormal GH polypeptide whose alteration causes a reduced somatogenic activity but enables it to react with anti-GH antibodies. Takahashi et al. identified a C > T transition in codon 77 that results in an Arg > Cys substitution in the GH1 gene of a subject diagnosed with bioinactive GH. The patient was heterozygous for the mutation and isoelectric focusing of his serum showed an abnormal GH peak in addition to a normal peak *(36)*. Surprisingly, his father was also heterozygous for the C > T transition but was of normal height and had normal isoelectric focusing results. The disparate findings in the father and his affected child were not explained.

3.5.5. GHRH Receptor Defects Heterogeneous mutations in the human GHRHR gene (OMIM 139191) have been identified. Most patients with these mutations have had poor growth since infancy and were extremely short. They failed to produce GH in response to standard provocative tests and had good responses to GH replacement. A nonsense mutation has been reported in the human GHRHR gene in two first cousins of a consanguineous Indian Moslem family with profound IGHD *(37)*. Both cousins were homozygous for a G > T transversion in exon 3, which converted a Glu to Ter (Glu27Ter) in their GHRHR genes. Subsequently, the same mutation has been identified in an isolate from the Indus valley of Pakistan *(38)*. A second mutation has been identified in a large isolate from Brazil diagnosed with an autosomal recessive form of IGHD. Affected subjects were found to be homozygous for a G > A transition of the first base of IVS1, which is predicted to alter splicing and results in an inactive protein product *(39)*.

3.5.6. Growth Hormone Resistance To be biologically active, GH (OMIM 139250) must bind to its transmembrane receptor GHR (OMIM 600946). GHR then dimerizes, activating intracellular signal transduction pathways resulting in the synthesis and secretion of IGF1 (OMIM 147440). IGF1 binds to its receptor (IGF1R, OMIM 147370) and activates its own signal transduction pathways, resulting in mitogenic and anabolic responses that lead to growth. Disruptions in GHR or IGF1 can cause GH resistance characterized by phenotypic features of GH deficiency associated with normal or high GH levels.

3.5.5.1. Laron Dwarfism I: Growth Hormone Receptor Defects Laron dwarfism is an autosomal recessive disorder caused by GH resistance as a result of defects in GHR (OMIM 600946). Although at the clinical level, Laron syndrome cases might be indistinguishable from GHD cases, they differ at the biochemical level. Patients with Laron dwarfism have low levels of IGF1, despite normal or increased levels of GH. This contrasts with the low levels of both IGF1 and GH that are seen in GHD. Importantly, exogenous GH does not induce an IGF1 response or restore normal growth in Laron dwarfism I cases because of GHR dysfunction. Although the majority of reported patients are Jewish, the disorder has been described in other ethnic groups. Laron dwarfs have the clinical appearance of severe IGHD with very delayed growth, abnormal facial appearance, high-pitched voice, and small male genitalia *(41–43)*. Length at birth might be short in relation to the birth weight, and tooth eruption and fontanelle closure are delayed. Laron dwarfs have the truncal obesity and the increased upper–lower segment body ratios typical of pituitary dwarfs. Although spontaneous hypoglycemia can occur, the production of other anterior pituitary hormones (ACTH, TSH, and gonadotropins) remains intact. Fasting GH levels are usually increased and range from normal to greater than 100 ng/mL. Plasma IGF-I levels are low and, in contrast to those of GH-deficient subjects, do not respond to exogenous rhGH *(41,42)*.

Although plasma levels of the GH-binding proteins (GHBP) that are derived from the extracellular domain of GHR are usually low in Laron dwarfism I cases, Woods et al. reported a homozygous point mutation in the intracellular domain of the GHR that caused Laron syndrome with elevated GHBP levels *(44)*. They predicted that the mutant GHR would not be anchored in the cell membrane but would be measurable in the serum as GHBP, thus explaining the phenotype of severe GH resistance combined with elevated circulating GHBP. Studies of the GHR genes of Laron dwarfism I cases have identified a variety of exon deletions and base substitutions. Although treatment with exogenous GH is ineffective in those with GHR dysfunction, replacement therapy with recombinant IGF1 has been shown to be effective.

Studies indicate that GH produced by Laron dwarfs reacts normally in radioreceptor assays with the GH receptors of normal hepatic cells *(42)*. This, along with their lack of response to exogenous GH, suggests the primary defect might be an abnormality of membrane receptors for GH. Molecular analysis of GH receptor (GHR) genes has shown deletions, point mutations, and splicing defects.

3.5.5.2. GHR Deletions The first examples of GHR mutations reported were deletions of portions of the gene encoding the extracellular domain *(45)*. Southern blotting showed altered restriction patterns of the GHR genes from two of nine patients with Laron syndrome who had no detectable GH-binding protein (GHBP) and very low levels of IGF-I. Although these and other studies were interpreted as showing deletions of exons 3, 5, 6, and part of 4 from the GHR gene, the mechanism by which these two noncontiguous deletions arose remains unclear.

3.5.5.3. GHR Point Mutations Over 50 different mutations in the GHR gene have been reported (see Human Gene Mutation Database, www.uwcm.ac.uk/uwcm/mg/search/ 119984.html). For example, Amselem et al. detected a T > C transition that converts the 96th residue of the extracellular domain from phenylalanine to serine *(46)*. Duquesnoy et al. demonstrated that cells transfected with this mutant cDNA lacked GH-binding activity *(47)*.

There are multiple different stop codon mutations of GHR genes in Laron dwarf patients *(48)*. For example, in a patient of northern European origin, a TGC (Cys) to TGA (stop) mutation was detected at codon 38 in exon 4, and a CGA (Arg) to TGA (stop) mutation was found at codon 43 in exon 4 of two

Mediterranean patients who were products of consanguineous marriages. Both stop codons truncate the GHR protein and delete most of its GHBP domain and all of its transmembrane and intracellular domains. These findings are consistent with the lack of GHBP in each of the patients with Laron syndrome. The mechanism of the CGA to TGA mutation is consistent with deamination of 5-methylcytosine that preferentially occurs in CpG dinucleotides. Such dinucleotides often represent "hot spots" for CG to TG or CG to CA mutations, and 17 occur within the GHR gene. Two of these, at nucleotides 181 and 703, occur in CGA codons that could yield stop codons *(46–49)*. A GAT(Asp) to CAT(His) mutation was identified in exon 6 of two unrelated kindreds with growth deficiency but normal GHBP levels. Conversion of the highly conserved aspartic acid residue to histidine was shown to prevent dimerization of the GH receptor which is necessary for GH action *(50,51)*.

3.5.5.4. GHR Splicing Defects Rosenbloom et al. identified 20 patients with Laron syndrome in an inbred population of Spanish extraction in southern Ecuador *(52)*. These patients were –6.7 to –10 SD below the mean height and had limited elbow extension, blue sclera, short limbs, hip degeneration, acrohypoplasia, and normal or superior intelligence. To determine the associated defect in the GHR gene, Berg et al. used denaturing gradient gel electrophoresis to analyze each exon of the GHR gene *(53)*. Unusual fragments derived from exon 6 showed abnormal mobility and DNA sequencing showed an A > G transition in the third position of codon 180, which is 24 nucleotides from the 3′ end of exon 6. Although this mutation does not cause an amino acid substitution, it produces a consensus 5′ SS sequence within exon 6. The resulting near-consensus 5′ SS within exon 6 causes aberrant splicing and deletion of eight amino acids of the 3′ end of exon 6. Deletion of these residues is thought to reduce the function of the GHR molecule *(53)*.

Two cousins of Pakistani descent were found to have a G > C transversion at the 5′ SS of exon 8 resulting in exon skipping. The mutant protein lacks the transmembrane and intracellular domains and results in elevated circulating levels of GH-binding protein in the affected patients *(44)*.

3.5.5.5. Laron Dwarfism II: Post GHR Defects Laron Dwarfism II is caused by post-GHR defects (see OMIM 245590). Patients with Laron dwarfism II have elevated serum GH, normal GHBP levels, and respond well to treatment with IGF-1, indicating their growth deficiency is the result of a post-GHR defect. Woods et al. described a patient with severe growth failure, sensorineural deafness, and mental retardation who was homozygous for a partial deletion of the IGF1 gene (54). RT-PCR analysis confirmed the deletion of exons 4–5 that would result in a severely truncated mature IGF1 peptide. Interestingly, this patient had only a slightly delayed bone age, suggesting that GH might directly stimulate bone maturation.

3.5.6. Combined Pituitary Hormone Deficiency Patients with CPHD vary in their clinical findings because they have deficiencies of one or more of the other pituitary trophic hormones (ACTH, FSH, LH, PRL, or TSH) in addition to GHD (OMIM 262600). Although most cases are sporadic, familial forms of CPHD can have autosomal recessive, autosomal dominant, or X-linked modes of inheritance. The clinical features of familial CPHD are identical to those of cases of nongenetic etiology. The phenotype varies with the specific trophic hormone deficiencies, which occur in decreasing order: gonadotropins (FSH, LH) > ACTH > TSH. Associated gonadotropin deficiency causes sexual immaturity and primary ammenorrhea in females and small external genitalia and lack of beard growth in males. TSH deficiency might become severe after GH replacement and ACTH deficiency might contribute to recurrent hypoglycemia. The severity of deficiencies of various trophic hormones exhibits interfamilial and intrafamilial variability in the various types of CPHD. Furthermore, the GH secretory responses to GHRH infusions vary from deficient to normal in different related individuals from the same families *(55)*.

3.5.6.1. HESX1 Mutations HESX1 is a transcription factor expressed in the thickened layer of oral ectoderm that gives rise to Rathke's pouch, the primordium of the anterior pituitary. Downregulation of HESX1 coincides with the differentiation of pituitary specific cell types. Dattani et al. found a missense HESX1 mutation (ARG53CYS) in a homozygous state in a brother and sister with septo-optic dysplasia, agenesis of the corpus callosum, and CPHD (OMIM 182230) (56).

3.5.6.2. LHX3 Mutations Murine LHX3 mRNA accumulates in Rathke's pouch and is thought to be involved in differentiation of pituitary cells. Netchine et al. identified two families with CPHD (OMIM 262600) caused by mutations in the LHX3 gene *(57)*. The phenotype associated with these mutations included the following: (1) severe growth retardation, (2) complete deficiency of all but one of the anterior pituitary hormones (ACTH), (3) elevated and anteverted shoulders with a short neck associated with severe restriction of rotation of the cervical spine, and (4) an enlarged anterior pituitary. The authors concluded that LHX3 is required for the proper development of all anterior pituitary cell types except corticotropes and that the rigid cervical spine phenotype is consistent with a function of LHX3 in the proper development of extrapituitary structures as well.

3.5.6.3. PIT1 Mutations Defects in the PIT1 gene cause familial CPHD with a different phenotype (OMIM 173110). PIT1 is an anterior pituitary-specific transcription factor, that regulates the expression of GH, PRL, and TSH. PIT1 is also required for pituitary cellular differentiation and function. PIT1 has functional domains that enable transactivation of other genes including GH, PRL, and TSH on binding to these genes. At least six different PIT1 mutations causing autosomal recessive and two others causing autosomal dominant CPHD have been found in humans in a subtype of pan-hypopituitary dwarfism associated with GH, PRL, and TSH deficiency (see OMIM 173110).

Mutations of either the PIT1 or PROP1 genes can cause CPHD in humans. Both of these genes are members of the POU (Pit-1, Oct-1, and Unc-86) homeodomain family of transcription factors and they each play an important role in pituitary development. Pit-1 gene defects were first identified in Snell and Jackson dwarf mice, which lack somatotropes, lactotropes, and thyrotropes and have severe deficiencies of GH, PRL, and TSH *(58)*. The Pit-1 (or GHF-1) gene, whose product binds to and activates GH, PRL, and TSH promoters, was found to have a T > G substitution at codon 261, giving a TGG (tryptophan)

to TGT (cysteine) in the Snell mice. A gross rearrangement of the Pit-1 gene was seen in the Jackson dwarf mouse. The deficiencies seen in these mice differ from most humans with CPHD, who have GH, TSH, gonadotropin, and ACTH deficiencies but increased PRL. In examining a subset of patients with GH, PRL, and TSH deficiency, eight different PIT1 mutations have been found in humans.

3.5.6.4. Autosomal Recessive PIT1 Mutations The first PIT1 mutation reported to cause CPHD was a T > C transition in codon 172, which changes a CGA (Arg) to TGA (Ter) *(59)*. The second PIT1 mutation reported was an A > G transition in codon 143, changing a CGA (Arg) to CAA (Gln) *(60)*. Both mutations occur in the POU-specific domain of PIT1 and are thought to affect binding of the PIT1 protein to DNA. The third and fourth PIT1 mutations were found in two Dutch families who had postnatal growth failure with complete deficiencies of GH. PRL and T4 levels fell after treatment with rhGH in one case and were low prior to treatment in the other *(61,62)*. One Dutch family had affected sibs whose T4 levels were initially normal. These sibs were homozygous for a C > G transversion in codon 158, which encoded a GCA (Ala) to CCA (Pro) substitution. This Ala158Pro mutation interferes with formation of PIT1 homodimers and greatly reduces transcription activation. The second Dutch family had affected sibs whose initial T4 levels were low. These sibs were genetic compounds with one deleted and one Ala158Pro PIT1 allele. Interestingly, this combination of defects was associated with more severe hypothyroidism and small anterior pituitary glands *(62)*. These cases emphasize the importance of determining T4 and PRL levels and TSH responses to TRH administration in those with CPHD. Because GH and TSH deficiency often occur together, failure of subjects to have PRL and TSH responses to TRH should raise the question of their having PIT1 gene defects. A fifth PIT1 mutation was identified in a Thai patient who was homozygous for a G to T transversion in codon 25 converting a GAA(Glu) to TAA (Ter) codon. This mutation resulted in complete loss of the POU homeodomain, which is necessary for DNA binding *(63)*. The sixth PIT1 mutation was found in a consanguineous family of Tunisian decent. All of the affected sibs were found to be homozygous for a T > G transversion in codon 135, converting a TTT (Phe) to TGT (Cys) in the POU-specific domain of PIT1 *(64)*.

3.5.6.5. Autosomal Dominant PIT1 Mutations Two dominant negative PIT1 mutations have been reported. Although the mechanism of action is not completely understood, neither of these mutations appear to inhibit binding of the mutant PIT1 protein to its target DNA. The first mutation, located in the POU homeodomain, is a C > T transition in codon 271 converting a CGG (Arg) to TGG (Trp) *(65)*. Three unrelated patients (two adults and one infant) have been reported to be heterozygous for this Arg271Trp mutation. Both adults had pituitary hypoplasia and the 2-mo-old had a normal pituitary by imaging. These findings suggest that PIT1 might be necessary for anterior pituitary survival and that the 2-mo-old will develop pituitary hypoplasia with age. The second dominant negative mutation was a C > T transition of the 24th codon converting a CCT (Pro) to CTT (Leu). This mutation resides within the major transactivating domain of PIT1, which is highly conserved in different species *(60)*.

3.5.6.6. PROP1 Mutations PROP1 or Prophet of PIT1 is a pituitary-specific homeodomain factor that is required for development of somatotropes, lactotropes, and thyrotropes of the anterior pituitary and for expression of PIT1. Multiple PROP1 gene mutations cause an autosomal recessive CPHD that has a different phenotype in humans (OMIM 601538). In addition to the deficiencies of GH, PRL, and TSH seen in those with PIT1 defects, subjects with PROP1 defects also have deficiencies of LH and FSH, which prevent the onset of spontaneous puberty. In some cases, ACTH deficiency develops in later life. The various PROP1 mutations include (1) a C > T transition in codon 120, which encoded a TGC (Arg) to CGC (Cys) substitution, (2) a T > A transversion that encodes a TTC (Phe) to ATC (Ile) substitution at codon 117, and (3) 2-bp AG deletion in codon 101 (101delAG) that causes a frameshift and results in a premature stop at codon 109. The resulting protein products from all three of these different PROP1 mutations have greatly reduced DNA binding and transactivation abilities *(66)*. 101delAG is a recurring mutation that is estimated to occur in about 55% of familial and 12% of sporadic CPHD cases *(67)*. A fourth PROP1 mutation is a 2-bp GA deletion in codon 51 (51delGA) *(68)*. Like the 101delAG mutation, the 51delGA mutation causes a frameshift that results in a premature stop codon. This mutation was found in 12% of familial and 21% of sporadic CPHD cases.

3.5.6.7. X-Linked CPHD Lagerstrom-Fermer et al. reported a family that included affected males suffering from variable degrees of CPHD (OMIM 312000) *(69)*. Some affected males died during the first day of life and had postmortem findings of hypoadrenalism, presumed to be the result of CPHD. Others had variable combinations of hypothyroidism, delayed pubertal development, and short stature because of GHD. All surviving patients exhibited mild to moderate mental retardation. They found linkage with markers in the Xq25-q26 region. Furthermore, they found an apparent extra copy of the marker DXS102 in affected males and heterozygous carrier females, suggesting that a segment including this marker was duplicated.

4. THE THYROID

4.1. DYSHORMONOGENESIS A variety of genetic defects could affect thyroid hormone synthesis, resulting in congenital hypothyroidism. Inability to transport iodide is caused by defects in the sodium–iodide symporter (NIS) (OMIM 601843) *(70)*. In Pendred syndrome, characterized by congenital hypothyroidism with goiter and deafness, the defect lies in the step of organification of iodide, caused by mutations in the SLC26A4 gene (OMIM 605646) *(71)*. Organification defects are also caused by mutations in thyroid peroxidase (TPO) (OMIM 606765) *(72)*.

4.2. THYROID HORMONE RESISTANCE Generalized thyroid hormone resistance is characterized by a reduced effect of T4 and T3 because of mutations in the thyroid hormone receptor-β (THRβ) (OMIM 190160). Patients typically have elevated levels of T4 and T3 with an inappropriately normal TSH. Most patients are clinically euthyroid, although some could have symptoms of hypothyroidism because of the lack of thyroid hormone effect. The majority of patients have autosomal

dominant-negative inheritance because mutations affect ligand binding or interaction with cofactors while preserving DNA-binding activity. The defective receptors occupy DNA-binding sites but are unable to activate transcription; their presence on the binding site prevents action of the normal allele *(73)*.

5. PARATHYROID DISORDERS

5.1. ALBRIGHT'S HEREDITARY OSTEODYSTROPHY

Albright's hereditary osteodystrophy (AHO), also called pseudohypoparathyroidism, is a clinical syndrome of short stature and short fourth and fifth metacarpals associated with hypocalcemia and ectopic calcification. It is caused by inactivating mutations in the GNAS1 gene (OMIM 139320), which leads to impaired parathyroid hormone action. The GNAS1 gene is paternally imprinted, meaning that normally only the maternal copy is expressed. Patients with AHO exhibit variable severity resulting from a combined effect of tissue-specific imprinting and haploinsufficiency. Thus, patients who inherit an abnormal GNAS1 gene from their mother are more severely affected (the normal paternal allele is imprinted and suppressed), and patients who inherit the gene from their father are less severely affected (the abnormal paternal allele is suppressed, and the normal maternal allele is expressed). Figure 2 shows an example pedigree showing the effect of imprinting on the severity of the clinical expression of AHO *(74,75)*.

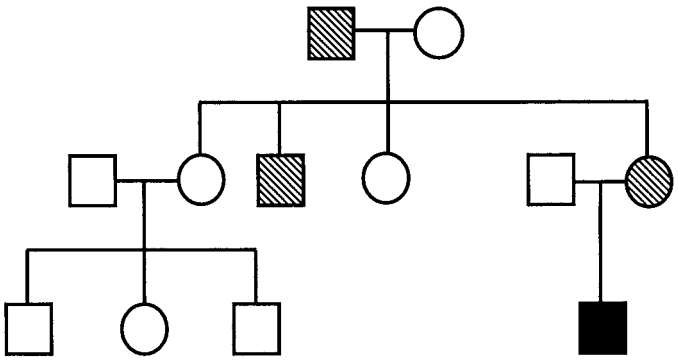


Fig. 2. Pedigree of a family with Albright hereditary osteodystrophy, demonstrating the effect of paternal imprinting. Hatched squares represent patients mildly affected because of silencing of the imprinted, mutated paternal allele; the black square represents a patient severely affected because of full expression of the nonimprinted, mutated maternal GNAS1 allele and silencing of the imprinted, normal paternal GNAS1 allele.

5.2. CALCIUM-SENSING RECEPTOR MUTATIONS

The calcium-sensing receptor (CASR) (OMIM 601199) is a plasma membrane G-protein-coupled receptor expressed in the parathyroid glands and kidneys *(76)*. Inactivating mutations in this receptor result in decreased calcium sensing with an altered "set point" for calcium, causing mild hypercalcemia *(77)*. This autosomal dominant syndrome is called familial hypocalciuric hypercalcemia (FHH). Infants with homozygous-inactivating mutations have neonatal severe hyperparathyroidism (NSPT), a more severe disorder with marked hypercalcemia *(77)*. Activating mutations have been described and lead to autosomal dominant hypocalcemia *(78,79)*.

6. LOSS OF HETEROZYGOSITY AS A MECHANISM OF ENDOCRINE DISEASE: THE MULTIPLE ENDOCRINE NEOPLASIA SYNDROMES

6.1. MULTIPLE ENDOCRINE NEOPLASIA TYPE 1

Multiple endocrine neoplasia type 1 (MEN1) is an autosomal dominant trait with marked intrafamilial variability, characterized by multiple tumors or hyperplasia of endocrine glands (see OMIM 131100). MEN1 is defined by the presence of two of the following three endocrine tumors: parathyroid adenoma or hyperplasia, entero-pancreatic endocrine tumors, and pituitary adenomas. Other possible manifestations include carcinoid tumors, adrenal adenomas, and lipomas. The MEN1 gene was identified through analysis of large kindreds with MEN1. It is located on chromosome 11q13 *(80)*. The gene product, menin, appears to function as a tumor suppressor gene. Loss of one copy of the gene (loss of heterozygosity) because of germline mutations leaves cells vulnerable to loss of the second allele through somatic mutations. Loss of the second allele, in turn, leads to tumor formation in affected tissues.

Commercial testing for MEN1 mutations will identify approx 80–90% of mutations. At-risk relatives of index patients should be tested. Identification of carrier status can identify those patients who require regular screening for the manifestations of MEN1 and exclude noncarriers from screening. In families whose mutation cannot be identified, 11q13 haplotype testing could identify carriers *(81)*.

6.2. MULTIPLE ENDOCRINE NEOPLASIA TYPES 2A AND 2B AND FAMILIAL MEDULLARY THYROID CARCINOMA (MEN2A, MEN2B, AND FMTC)

Multiple endocrine neoplasia type 2A (MEN2A) is an autosomal dominant syndrome defined by medullary thyroid carcinoma, pheochromocytoma, and parathyroid adenomas (OMIM 171400). Ninety percent of MEN2A patients will have medullary thyroid carcinoma, 50% will have pheochromocytomas and 20–30% will have parathyroid tumors of MEN2B consists of medullary thyroid carcinoma and pheochromocytomas in association with marfanoid body habitus and intestinal and mucosal ganglioneuromatosis. Familial medullary thyroid carcinoma (FMTC) can exist in isolation. Hirschsprung's disease can be seen in association with either MEN2A or FMTC.

All three disorders (MEN2A, MEN2B, and FMTC) are caused by mutations in the RET gene (OMIM 164761). RET is located on chromosome 10 and encodes a membrane-bound tyrosine kinase. A limited number of mutations are associated with the MEN2 syndromes, mainly in exons 10, 11, 13, 14, 15, and 16. Testing of these exons is commercially available and will detect the majority of mutations. Prior to the availability of RET testing, patients who were possible carriers were monitored for medullary thyroid carcinoma (MTC) by following basal and stimulated calcitonin levels; however, these have higher false-positive and false-negative rates than mutation testing *(81)*. Determining RET carrier status in children at risk is particularly important because prophylactic thyroidectomy might be life-saving. In one study, 67% of juvenile RET mutation carriers subsequently shown to have MTC or C-cell hyperplasia (a premalignant lesion) at thyroidectomy had normal

basal calcitonin levels *(82)*. MTC has been reported in infancy, highlighting the need for early testing and decisions about prophylactic thyroidectomy.

6.3. PATHOLOGIC ALLELIC VARIANTS OF THE RET GENE The major disease-causing mutations are nonconservative substitutions located in one of six cysteine codons in the extracellular domain of the encoded protein. They include codons 609, 611, 618, and 620 in exon 10 and codons 630 and 634 in exon 11 *(83)*. All of these variants have been identified in families with MEN2A, and some have been identified in families with FMTC. Mutations in these sites have been detected in 95% of MEN2A families *(84)*. Approximately 95% of all individuals with the MEN2B phenotype have a single-point mutation in the tyrosine kinase domain of the RET gene at codon 918 in exon 16, which substitutes a threonine for methionine *(85,86)*. A second point mutation at codon 883 has been found in four MEN2B patients *(87,88)*. In addition to the mutations of the cysteine residues in exons 10 and 11 that have been found in families with MEN 2A, mutations in codons 631, 768, 790, 791, 804, 844, and 891 have also been identified in a small number of families *(89–92)*. A mutation at codon 603 was reported in one family and appeared to be associated with both MTC and papillary thyroid cancer *(93)*. Duplication mutations have been reported in two families *(94,95)*. Rare families have two mutations in cis configuration; for example, alteration of both codons 634 and 635 in one family with MEN2A *(96)*, alteration of both codons 804 and 844 in one family with FMTC *(97),* and alteration of codons 804 and 806 in an individual with MEN2B *(98)*.

6.4. MOLECULAR GENETIC TESTING IN MEN SYNDROMES The MEN syndromes are a group of conditions in which molecular genetic testing guides therapy and can be life-saving. Testing for MEN1 or RET mutations can confirm the diagnosis in affected patients and, in the case of RET mutations, guide optimal timing of thyroidectomy. For unaffected patients at risk, testing can identify carriers who will need life-long surveillance for the manifestations of these syndromes and can exclude noncarriers from costly and invasive testing.

6.4.1. Resources for Genetic Testing Molecular diagnostic tests are increasingly available through commercial laboratories. Testing for conditions such as MEN2A and MEN2B is often performed in a two-tiered manner, with initial testing for the most common variants followed by screening of the whole gene only if the most common variants are not present. For some conditions, testing is available through research laboratories only. Websites such as GeneTests–GeneClinics (http://www.geneclinics.org/) provide listings of clinical and research laboratories that offer molecular genetic testing. For example, at the time of this writing, there are 3 clinical laboratories offering genetic testing and 2 research laboratories soliciting samples for MEN1, and 12 clinical labs offering RET testing and 1 research lab soliciting samples for MEN2. Referral to a geneticist or genetic counselor might be useful in obtaining assistance in diagnosis and management and in counseling the patient and family members regarding issues such as risk of recurrence of a disorder, reproductive decisions, or testing of additional family members.

6.4.2. Genetic Testing in Children Genetic testing in children presents special considerations. In general, genetic testing in children should only be considered when the results would make an immediate and important change in the patient's management. For example, a young child with a positive test for an MEN2B-associated RET mutation would be referred for thyroidectomy. Predictive testing for adult-onset diseases such as Huntington's chorea or carrier testing for diseases such as Tay–Sachs disease is generally not recommended until the child is old enough to decide for himself or herself. Prior to performing any genetic testing in children, informed consent from the parent or legal guardian and assent from the child should be obtained. For further discussion of issues related to genetic testing in children, the reader is referred to the published statements of the American Academy of Pediatrics and the American Society of Human Genetics *(99,100)*.

7. THE MOLECULAR ENDOCRINOLOGY OF DIABETES MELLITUS

Diabetes mellitus is a metabolic syndrome characterized by hyperglycemia resulting from variable defects in insulin secretion and action. Traditionally, diabetes mellitus has been divided into four categories: Type 1 diabetes, characterized by absolute insulin deficiency and requirement of insulin therapy to sustain life; Type 2 diabetes, characterized by variable defects in insulin secretion and action, without an absolute need for insulin therapy (although insulin might be necessary for adequate control of hyperglycemia); "other specific types" of diabetes; and gestational diabetes. Type 1 diabetes is further subdivided into Type 1A, or autoimmune, diabetes and Type 1B, or idiopathic diabetes. The "other" category is very large, encompassing diabetes resulting from other illness or medications (such as cystic fibrosis-related diabetes) as well as many genetic syndromes that include diabetes *(101)*. It is becoming increasingly apparent, however, that these categories are artificial and that the clinical syndrome of "diabetes mellitus" represents a heterogeneous group of disorders with widely varying etiologies.

7.1. MATURITY ONSET DIABETES OF THE YOUNG Maturity-onset diabetes of the young (MODY) is a syndrome of non-insulin-dependent diabetes characterized by autosomal dominant inheritance with at least two affected family members in at least two generations, at least one diagnosed before the age of 25. MODY patients are typically not obese (unlike T2DM patients) and have normal insulin sensitivity as well as negative islet cell autoantibodies *(102–104)*. To date, six genes have been identified as causes of MODY (Table 2). Of these, one (glucokinase) is an enzyme, and the rest are transcription factors that are important in pancreatic and islet cell development as well as in maintenance of insulin secretion. These disorders are important because they offer insight into the regulatory pathways and because they suggest possible pathogenic mechanisms for other forms of diabetes. In addition, molecular diagnostic testing for two of the most common forms of MODY is now available.

MODY2 is caused by defects in the enzyme glucokinase (GCK) (OMIM 138079) and is the second most common form of MODY in European populations *(105)*. GCK catalyzes the phosphorylation of glucose to glucose-6-phosphate, the

rate-limiting step in the glycolytic pathway. This allows GCK to serve as the glucose sensor for the insulin-secreting β cells of the pancreatic islets *(106)*. More than 130 mutations causing MODY2 have been described since the original description in 1992 *(107,108)*. These mutations result in decreased activity of the enzyme, which, in turn, leads to decreased glucose sensing. The net result is an increase in the glucose "set point" for insulin secretion. Patients with MODY2/GCK deficiency have mild hyperglycemia both fasting and postprandially. This hyperglycemia often does not require therapy and diabetes-associated complications are rare *(103)*. The remaining forms of MODY are due to mutations in several transcription factors (see Table 2 and below). These forms of MODY are generally more severe than MODY2/GCK, and patients may require treatment with oral agents and/or insulin. Patients with these forms of diabetes are at risk for diabetes-related complications.

MODY1 is caused by mutations in hepatic nuclear factor 4α (HNF4α) (OMIM 600281), a transcription factor originally identified from rat liver *(109,110)*. HNF4α and its relatives are now understood to be expressed in many tissues and are involved in a transcriptional regulation network that controls both pancreatic development and insulin secretion *(110)*. In particular, HNF4α has been found to have a pancreas-specific promoter (designated P2) located 46 kb upstream to the P1 promoter, which is expressed in hepatocytes *(111)*. This upstream promoter contains binding sites for HNF1α, HNF1β, and IPF1, all of which are also associated with MODY syndromes *(111)*. MODY1 is most commonly caused by mutations in the HNF4α coding region, which lead to decreased DNA binding and/or decreased transactivation *(110)*; however, one family has been reported in which the mutation lies in the IPF-1-binding site of the P2 promoter *(111)*, and another family has been reported with a balanced translocation {karyotype 46,XX t(3;20)(p21.2;q12)} with a breakpoint between the P2 promoter and the rest of the HNF4α gene *(112)*.

MODY3 is the most common variant of MODY in European populations *(105)* and is caused by mutations in the hepatic nuclear factor 1α (HNF1α) (OMIM 142410) *(113)*. Mutations have been described throughout the HNF1α gene and its promoter, and there appears to be a mutation "hot spot" in exon 4 *(114)*.

MODY4, caused by mutations in the insulin promoter factor 1 (IPF1) (OMIM 600733) gene, is a rare form of MODY *(115)*. IPF1 is a transcription factor that is essential for normal pancreas and islet cell development as well as islet cell function in the adult *(116)*. Heterozygous dysfunction of the gene results in a MODY syndrome similar to the other forms of transcription factor-related MODY, whereas homozygotes have pancreatic agenesis and permanent insulin-requiring diabetes from the neonatal period *(115,117)*.

MODY5 is the result of mutations in the hepatocyte nuclear factor-1β gene (HNF1β) (OMIM 189907) *(118)*. Mutations in HNF1β are associated with progressive, nondiabetic renal failure and renal cysts in addition to diabetes *(118,119)*, and one mutation has also been associated with vaginal and uterine aplasia in female carriers *(120)*.

The most recent gene to be associated with a MODY syndrome (MODY6) is neurogenic differentiation 1 (NEUROD1, also called BETA2) (OMIM 601724) *(121)*. NEUROD1 is a transcription factor that regulates insulin gene transcription; mice that are homozygous null for NEUROD1 have abnormal pancreatic islets, suggesting a role in pancreatic development as well *(122)*.

7.2. OTHER MONOGENIC DIABETES SYNDROMES

Wolcott–Rallison syndrome (OMIM 226980) is an autosomal recessive disorder characterized by neonatal-onset diabetes in association with multiple epiphyseal dysplasia, decreased growth, osteoporosis, and renal insufficiency. The causative gene has been identified as eukaryotic translation initiation factor 2-α kinase 3 (EIF2AK3) (OMIM 604032) *(123)*. This kinase is thought to play a role in regulating protein flux through the endoplasmic reticulum (ER), in particular in coupling the rate of insulin synthesis with the capacity for peptide processing in the ER *(124)*.

Wolfram syndrome (WFS) (OMIM 222300) is also an autosomal recessive disorder, characterized by the mnemonic DIDMOAD (diabetes insipidus, diabetes mellitus, optic atrophy, and deafness). Diabetes typically occurs in childhood. The causative gene, WFS1 (OMIM 606201), has been identified (125,126); its protein product, wolframin, is a transmembrane protein of uncertain function, although it has been hypothesized to play a role in β-cell and neuronal cell survival pathways based on pathological data. Mutations have been found in more than 90% of patients with WFS and are located throughout the gene with a concentration in exon 8 *(127)*.

The syndrome of maternally transmitted diabetes and deafness (OMIM 520000) is caused by a mutation (3243A > G) in the mitochondrial tRNA-LEU1 gene (MTTL1) (OMIM 590050) *(128)*. Diabetes in these patients is similar to Type 2 diabetes, and defects of glucose-regulated insulin secretion have been found *(129)*.

7.3. THE GENETICS OF TYPE 1A (AUTOIMMUNE) DIABETES MELLITUS

Type 1A diabetes mellitus (T1ADM) is a relatively common disease, affecting approximately 1 in 300 children *(130)*. It is caused by autoimmune destruction of the pancreatic islets. Twin studies reveal a clear genetic predisposition, with concordance rates of 21–70% in monozygotic twins and 0–13% in dizygotic twins. The major source of this genetic predisposition lies in the human leukocyte antigen (HLA) locus, with known haplotypes that confer either high risk for or protection from the disease (for review, see ref. *130*). However, there appear to be modifying genes and environmental factors. One such modifying locus is the IDDM2 locus on chromosome 11p15, which corresponds to a variable number of tandem repeats (VNTR) 5′ to the insulin gene. These insulin VNTRs are divided into classes based on the number of repeats: Class I (26–63), Class II (approx 80 repeats), and Class III (140–200 repeats). The number of repeats correlates with insulin gene expression in the thymus, and lower numbers of repeats create a higher risk of diabetes, whereas higher numbers are protective. Higher insulin gene expression in the thymus might allow T-cells that react to self-antigen to be selected out, thus decreasing the potential for autoimmunity *(130)*.

There are two monogenic syndromes associated with T1ADM. The first of these is known as autoimmune polyendocrinopathy syndrome type 1 (APS1), or autoimmune

polyendocrinopathy–candidiasis–ectodermal dystrophy (APECED). The most characteristic features of this syndrome are chronic mucocutaneous candidiasis, hypoparathyroidism, Addison's disease, and T1ADM; other features include other autoimmune endocrinopathies (such as hypothyroidism or hypogonadism), malabsorption syndromes, pernicious anemia, and alopecia. The gene for this disorder is the autoimmune regulator or AIRE (OMIM 607358), which appears to be a transcription factor. The exact mechanism by which it causes the syndrome is not known *(131,132)*.

The second monogenic syndrome is X-linked autoimmunity–allergic dysregulation syndrome (XLADD, also called immunodysregulation, polyendocrinopathy, and enteropathy, X-linked or IPEX) (OMIM 304790), a syndrome of multisystem autoimmunity including diabetes and thyroid disease as well as frequent infections, diarrhea, and hemolytic anemia. The disorder is usually fatal. Mutations in the transcription factor forkhead box P3 gene (FOXP3) (OMIM 300292) are causative *(133)*.

7.4. TYPE 2 DIABETES: A POLYGENIC FORM OF DIABETES Type 2 diabetes mellitus (T2DM) arises out of a combination of insulin resistance and insulin secretory defects. This interaction is, in turn, influenced by environmental factors such as diet, obesity, and physical activity. T2DM is known to have an even stronger genetic predisposition than T1DM, with a high concordance rate among identical twins (approx 80% as opposed to dizygotic twins, for whom the risk is approx 10%) and an increased incidence in first-degree relatives (approx 10%) *(134,135)*. Genome scanning approaches have shown linkages to a large number of chromosomal locations *(136)*, which vary among populations. For only one of these, 2q37 in Mexican-Americans, has a candidate gene been localized and shown to associate with diabetes. That gene is calpain-10 (CAPN10) (OMIM 605286), a type of protease *(137)*; how genetic variations in it lead to diabetes is not well understood, and association of T2DM with CAPN10 has not been reproduced in other populations *(138)*.

The alternative approach to identifying genes associated with T2DM is to screen populations of patients with T2DM for variations in plausible candidate genes. At least 17 genes have been studied *(139)*; such association studies are often not reproducible and associations might vary from population to population. However, at least three genes (ABCC8, PPARG, and SLC2A1) are significantly associated with T2DM when results from multiple studies are pooled *(140)*.

It appears likely that susceptibility to T2DM is the result of combinations of variations in many genes. Although each variation in itself would be insufficient to cause diabetes, combinations of variations would create risk haplotypes, which would then lead to diabetes under environmental influences such as diet.

8. CONCLUSION

Our understanding of the molecular basis of endocrine diseases is continually evolving. As new molecular mechanisms are elucidated, it is expected that these advances will eventually translate to the realm of molecular diagnostic testing. In the near future, molecular testing will come into increasing use in the clinic, with significant implications for future patient care.

REFERENCES

1. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921, 2001.
2. Ito, M., Mori, Y., Oiso, Y., et al. A single base substitution in the coding region for neurophysin II associated with familial central diabetes insipidus. *J. Clin. Invest.* **87:**725–728, 1991.
3. Kadowaki, T., Bevins, C. L., Ojamaa, K., et al. Two mutant alleles of the insulin receptor gene in a patient with extreme insulin resistance. *Science* **240:**787–790, 1988.
4. Krook, A., Brueton, L., and O'Rahilly, S. Homozygous nonsense mutation in the insulin receptor gene in infant with leprechaunism. *Lancet* **342:**277–278, 1993.
5. Wertheimer, E., Lu, S. P., Backeljauw, P. F., et al. Homozygous deletion of the human insulin receptor gene results in leprechaunism. *Nat. Genet.* **5:**71–73,1993.
6. Hone, J., Accili, D., Psiachou, H., et al. Homozygosity for a null allele in a patient with leprechaunism. *Hum. Mutat.* **6:**17–22, 1995.
7. Longo, N., Wang, Y., Smith, S. A., et al. Genotype–phenotype correlation in inherited severe insulin resistance. *Hum. Mol. Genet.,***11:** 1465–1475, 2002.
8. Lubahn, D. B, Brown, T. R., Simental, J. A., et al. Sequence of the intron/exon junctions of the coding region of the human androgen receptor gene and identification of a point mutation in a family with complete androgen insensitivity. *Proc. Natl. Acad. Sci. USA* **86:**9534–9538, 1989.
9. McPhaul, M. J. Molecular defects of the androgen receptor. *Rec. Prog. Horm. Res.***57:**181–194, 2002.
10. Weinstein, L. S., Shenker, A., Gejman, P. V., et al. Activating mutations of the stimulatory G protein in the McCune–Albright syndrome. *N. Engl. J. Med.* **325:**1688–1695, 1991.
11. Schwindinger, W. F,, Francomano, C. A., and Levine, M. A. Identification of a mutation in the gene encoding the α subunit of the stimulatory G protein of adenyl cyclase in McCune–Albright syndrome. *Proc. Natl. Acad. Sci.USA* **89:**5152–5156, 1992.
12. Lumbroso, S., Paris, F., and Sultan, C. McCune–Albright syndrome: molecular genetics. *J. Pediatr. Endocrinol. Metab.* **15:**875–882, 2002.
13. Rimoin, D. L., Phillips, J.A. III. Genetic disorders of the pituitary gland, in *Principles and Practice of Medical Genetics*, 3rd ed. Rimoin, D. L., Connor, J. M., and Pyeritz, R. E., eds Churchill Livingstone, New York, 1997, pp. 1331–1364.
14. Cogan, J. D. and Phillips, J.A. III. Inherited defects in growth hormone synthesis and action, in *The Metabolic and Molecular Bases of Inherited Disease,* 8th ed., Scriver, C.R., Beaudet, A.L., Sly, W. S., and Valle, D., eds. McGraw–Hill, New York, 2001, pp. 4159–4180.
15. Vnencak–Jones, C. L., Phillips, J. A., III, Chen, E. Y., et al. Molecular basis of human growth hormone gene deletions. *Proc. Natl. Acad. Sci. USA* **85:**5615–5619, 1988.
16. Mullis, P. E., Akinci, A., Kanaka, C. H., et al. Prevalence of human growth hormone-1 gene deletions among patients with isolated growth hormone deficiency from different populations. *Pediatr. Res.* **31:**532–534, 1992.
17. Duquesnoy, P., Amselem, S., Gourmelen, M., et al. A frameshift mutation causing isolated growth hormone deficiency type 1A. *Am. J.Hum. Genet* **47:**A110, 1990.
18. Cogan, J. D., Phillips, J. A., III, Sakati, N. A., et al. Molecular analysis of familial growth hormone deficiency. *The Endocrine Society Program and Abstracts* 321, 1992.
19. Cogan, J. D., Phillips, J. A., III, Sakati, N. A., et al. Heterogeneous growth hormone (GH) gene mutations in familial GH deficiency. *J. Clin. Endocrinol. Metab.* **76:**1224–1228, 1993.
20. Phillips, J. A., III, Parks, J. S., Hjelle, B. L., et al. Genetic basis of familial isolated growth hormone deficiency type I. *J. Clin. Invest.* **70:**489–495, 1982.
21. Rimoin, D. L., and Schechter, J. E. Histological and ultrastructural studies of isolated growth hormone deficiency. *J. Clin. Endocrinol. Metab.***37:**725–735, 1973.
22. Rogol, A. D., Blizzard, M. R., Foley, T. P., Jr., et al. Growth hormone releasing hormone and growth hormone: genetic studies in familial growth hormone deficiency. *Pediatr. Res.* **19:**489–492, 1985.

23. Abdul-Latif, H. D., Brown, M. R., Parks, J. S., et al. Mutation of intron 4 of the GH1 gene causes GH deficiency. *The Endocrine Society Program and Abstracts,* 470, 1995.
24. Igarashi, Y., Ogawa, M., Kamijo, T., et al. A new mutation causing inherited growth hormone deficiency: a compound heterozygote of a 6.7 kb deletion and two base deletion in the third exon of the GH1 gene. *Hum. Mol. Genet.* **2:**1073–1074, 1993.
25. Rimoin, D. L. Genetic disorders of the pituitary gland, in Emery, A. E. H. and Rimoin, D. L., eds. Edinburgh, *Principles and Practice of Medical Genetics,* Churchill Livingstone, 1983, pp. 1134–1151.
26. McGuinness, L., Magoulas, C., Sesay, A. K., et al. Autosomal dominant growth hormone deficiency disrupts secretory vesicles in vitro and in vivo in transgenic mice. *Endocrinology* **144:**720–731, 2003.
27. Binder, G., and Ranke, M. B. Screening for growth hormone (GH) gene splice–site mutations in sporadic cases with severe isolated GH deficiency using ectopic transcript analysis. *J. Clin. Endocrinol. Metab.* **80:**1247–1252, 1995.
28. Cogan, J. D., Ramel, B., Lehto, M., et al. A recurring dominant-negative mutation causes autosomal dominant growth hormone deficiency. *J. Clin. Endocrinol.Metab.* **80:**3591–3595, 1995.
29. Cogan, J. D., Prince, M. A., Lekhakula, S., et al. A novel mechanism of aberrant pre-mRNA splicing in humans. *Hum. Mol. Genet.* **6:**909–912, 1997.
30. Moseley, C. T., Mullis, P. E., Prince, M. A., et al. An exon splice enhancer mutation causes autosomal dominant growth hormone deficiency. *J. Clin. Endocrinol. Metab.* **87:**847–852, 2002.
31. Wajnrajch, M. P., Gertner, J. M., Mullis, P. E., et al. Arg183His, a new mutational "hot-spot" in the growth hormone gene causing isolated GH deficiency type II. *J. Endocr. Genet.* **1:**125–135, 2001.
32. Duriez, B., Duquesnoy, P., Dastot, F., et al. An exon-skipping mutation in the btk gene of a patient with X-linked agammaglobulinemia and isolated growth hormone deficiency. *FEBS Lett.* **346:**165–170, 1994.
33. Laumonnier, F., Ronce, N., Hamel, B. C. J., et al. Transcription factor SOX3 is involved in X-linked mental retardation with growth hormone deficiency. *Am.J.Hum.Genet.* **71:**1450–1455, 2002.
34. Kowarski, A. A., Schneider, J., Ben-Galim, E., et al. Growth failure with normal serum RIA-GH and low somatomedin activity: Somatomedin restoration and growth acceleration after exogenous GH. *J. Clin. Endocrinol.Metab.* **47:**461–464, 1978.
35. Rudman, K., Kutner, M. H., Blackston, R. D., et al. Children with normal-variant short stature: treatment with human growth hormone for six months. *N. Engl. J. Med.* **305:**123–131, 1981.
36. Takahashi, Y., Kaji, H., Okimura, Y., et al. Short stature caused by a mutant growth hormone. *N.Engl. J. Med.* **334:**432–436, 1996.
37. Wajnrajch, M. P., Gertner, J. M., Harbison, M. D., et al. Nonsense mutation in the human growth hormone-releasing hormone receptor causes growth failure analogous to the little (lit) mouse. *Nat. Genet.* **12:**88–90, 1996.
38. Bauman, G., and Maheshwari, H. The dwarfs of Sindh: severe growth hormone deficiency caused by a mutation in the GH-releasing hormone receptor gene. *Acta Paediatr,* **423(Suppl.):**33–38, 1997.
39. Salvatori, R., Gondo, R. G., de Aguirar Oliveira, M. H., et al. Familial isolated growth hormone deficiency due to a novel mutation in the growth hormone-releasing hormone receptor. *J. Clin. Endocrinol.Metab.* **84:**917–923, 1999.
40. Pertzelan, A., Adam, A., and Laron, Z. Genetic aspects of pituitary dwarfism due to absence or biological inactivity of growth hormone. *Israel J.Med. Sci.* **4:**895–900, 1968.
41. Laron, Z., Pertzelan, A., Karp, M., et al. Administration of growth hormone to patients with familial dwarfism with high plasma immunoreactive growth hormone: measurement of sulfation factor, metabolic and linear growth response. *J. Clin. Endocrinol. Metab.* **33:**332–342, 1971.
42. Jacobs, L. S., Sneid, S. D., Garland, J. T., et al. Receptor-active growth hormone in Laron dwarfism. *J. Clin. Endocrinol. Metab.* **42:** 403–406, 1976.
43 Adam, A., Josefsberg, Z., Pertzelan, A., et al. Occurrence of four types of growth hormone-related dwarfism in Israeli communities. *Eur.J. Pediatr.* **137:**35–39, 1981.
44. Woods, K. A., Fraser, N. C., Postel-Vinay, M. C., et al. A homozygous splice site mutation affecting the intracellular domain of the growth hormone (GH) receptor resulting in Laron syndrome with elevated GH-binding protein. *J. Clin. Endocrinol. Metab.* **81:**1686–1690, 1996.
45. Godowski, P. J., Leung, D. W., Meacham, L. R., et al. Characterization of the human growth hormone receptor gene and demonstration of a partial gene deletion in two patients with Laron-type dwarfism. *Proc. Natl. Acad. Sci. USA* **86:**8083–8087, 1989.
46. Amselem, S., Duquesnoy, P., Attree, O., et al. Laron dwarfism and mutation of the growth hormone-receptor gene. *N. Engl. J. Med.* **321:**989–995, 1989.
47. Duquesnoy, P., Sobrier, M. L., Amselem, S., et al. Defective membrane expression of human growth hormone (GH) receptor causes Laron-type GH insensitivity syndrome. *Proc. Natl. Acad. Sci. USA* **88:**10,272–10,276, 1991.
48. Amselem ,S., Sobrier, M. L., Duquesnoy, P., et al. Recurrent nonsense mutations in the growth hormone receptor from patients with Laron dwarfism. *J.Clin. Invest.* **87:**1098–1102, 1991.
49. Phillips, J. A., III. Molecular biology of growth hormone receptor dysfunction. *Acta Paediatr. Scand.* **383(Suppl.):**127–131, 1992.
50. Buchanan, C. R., Maheshwari, H. G., Norman, M. R., et al. Laron-type dwarfism with apparently normal high affinity serum growth hormone-binding protein. *Clin. Endocrinol.* **35:**179–185, 1991.
51. Duquesnoy, P., Sobrier, M. L., Duriez, B., et al. A single amino acid substitution in the exoplasmic domain of the human growth hormone (GH) receptor confers familial GH resistance (Laron syndrome) with positive GH-binding activity by abolishing receptor homodimerization. *EMBO J.* **13:**1386–1395, 1994.
52. Rosenbloom, A. L., Guevara-Aguirre, J., Rosenfeld, R. G., et al. The little women of Loja—growth hormone-receptor deficiency in an inbred population of southern Ecuador. *N. Engl. J. Med.* **323:** 1367–1374, 1990.
53. Berg, M. A., Guevara-Aguirre, J., Rosenbloom, A. L., et al. Mutation creating a new splice site in the growth hormone receptor genes of 37 Ecuadorian patients with Laron syndrome. *Hum. Mutat.* **1:**24–32, 1992.
54. Woods, K. A., Camacho-Hubner, C., Savage, M. O., et al. Intrauterine growth retardation and postnatal growth failure associated with deletion of the insulin-like growth factor 1 gene. *N. Engl. J.Med.* **335:**1363–1367, 1996.
55. Rogol, A. D., Blizzard, M. R., Foley, T. P., Jr., et al. Growth hormone releasing hormone and growth hormone: genetic studies in familial growth hormone deficiency. *Pediatr. Res.* **19:**489–492, 1985.
56. Dattani, M. T., Martinez-Barbera, J. P., Thomas, P. Q., et al. Mutations in the homeobox gene HESX1/Hesx1 associated with septo-optic dysplasia in human and mouse. *Nat. Genet.* **19:**125–133, 1998.
57. Netchine, I., Sobrier, M. L., Krude, H., et al. Mutations in LHX3 result in a new syndrome revealed by combined pituitary hormone deficiency. *Nat. Genet.* **25:**182–186, 2000.
58. Li, S., Crenshaw, E. B., 3rd, Rawson, E. J., et al. Dwarf locus mutants lacking three pituitary cell types result from mutations in the POU-domain gene pit-1. *Nature* **347:**528–533, 1990.
59. Tatsumi, K., Miyai, K., Notomi, T., et al. Cretinism with combined hormone deficiency caused by a mutation in the Pit-1 gene. *Nat. Genet.* **1:**56–58, 1992.
60. Ohta, K., Nobukuni, Y., Mitsubichi, H., et al. Mutations in the Pit-1 gene in children with combined pituitary hormone deficiency. *Biochem. Biophys. Res. Commun.* **189:**851–855, 1992.
61. Wit, J. M., Drayer, N. M., Jansen, M., et al. Total deficiency of growth hormone and prolactin, and partial deficiency of thyroid stimulating hormone in two Dutch families: a new variant of hereditary pituitary deficiency. *Horm. Res.* **32:**170–177, 1989.
62. Pfaffl, R. W., DiMattia, G. E., Parks, J. S., et al. Mutation of the POU-specific domain of Pit-1 and hypopituitarism without pituitary hypoplasia. *Science* **257:**1118–1121, 1992.
63. Irie, Y., Tatsumi, K., Ogawa, M., et al. A novel E250X mutation of the Pit-1 gene in a patient with combined pituitary hormone deficiency. *Endocr. J.* **42:**351–354, 1995.

64. Pelligrini-Bouiller, I., Belicar, P., Barlier, A., et al. A new mutation of the gene encoding the transcription factor Pit-1 is responsible for combined pituitary hormone deficiency. *J. Clin. Endocrinol. Metab.* **81:**2790–2796, 1996.
65. Radovick, S., Nations, M., Du, Y., et al. A mutation in the POU-homeodomain of Pit-1 responsible for combined pituitary hormone deficiency. *Science* **257:**1115–1118, 1992.
66. Wu, W., Cogan, J. D., Pfaffle, R. W., et al. Mutations in PROP1 cause familial combined pituitary hormone deficiency. *Nat. Genet.* **18:**147–149, 1998.
67. Cogan, J. D., Wu, W., Phillips, J. A., III, et al. The PROP1 2-bp deletion is a common cause of CPHD. *J. Clin. Endocrinol. Metab.* **83:**3346–3349, 1998.
68. Fofanova, O., Takamura, N., Kinoshita, E., et al. Compound heterozygous deletion of the PROP-1 gene in children with combined pituitary hormone deficiency. *J. Clin. Endocrinol. Metab.* **83:** 2601–2604, 1998.
69. Lagerstrom-Fermer, M., Sundvall, M., Johnsen, E., et al. X-linked recessive panhypopituitarism associated with a regional duplication in Xq25-q26. *Am. J. Hum. Genet.* **60:**910–916, 1997.
70. Fujiwara, H., Tatsumi, K., Miki, K., et al. Congenital hypothyroidism caused by a mutation in the Na+/I- symporter. *Nat. Genet.* **16:**124–125, 1997.
71. Online Mendelian Inheritance in Man, OMIM (TM). Johns Hopkins University, Baltimore, MD. MIM Number: {605646}: {12/4/2002}. Available at http://www.ncbi.nlm.nih.gov/omim/ (accessed December 13, 2002).
72. Online Mendelian Inheritance in Man, OMIM (TM). Johns Hopkins University, Baltimore, MD. MIM Number: {606765}. Available at http://www.ncbi.nlm.nih.gov/omim/ (accessed March, 30, 2003).
73. Refetoff, S., Weiss, R. E., Usala, S. J., et al. The syndromes of resistance to thyroid hormone. *Endocr. Rev.* **14:**348–399, 1993.
74. Levine, M. A., Ahn, T. G. , Klupt, S. F., et al. Genetic deficiency of the alpha subunit of the guanine nucleotide-binding protein Gs as the molecular basis for Albright hereditary osteodystropy. *Proc. Natl. Acad. Sci. USA* **85:**617–621, 1998.
75. Farfel, Z., Bourne, H. R., Iiri, T. The expanding spectrum of G protein diseases. *N. Engl. J. Med.* **340:**1012–1020, 1999.
76. Online Mendelian Inheritance in Man, OMIM (TM). Johns Hopkins University, Baltimore, MD. MIM Number: {601199}. Available at http://www.ncbi.nlm.nih.gov/omim/ (accessed December 10, 2002).
77. Pollak, M. R., Brown, E. M., Chou, Y. W., et al. Mutations in the human Ca2+-sensing receptor gene cause familial hypocalciuric hypercalcemia and neonatal severe hyperparathyroidism. *Cell* **75:**1297–1303, 1993.
78. Pollak, M. R., Brown, E. M., Estep, H. L., et al. Autosomal dominant hypocalcemia caused by a Ca2+-sensing receptor gene mutation. *Nat. Genet.* **8:**303–307, 1994.
79. Perry, Y. M., Finegold, D. N., Armitage, M. M., et al. A missense mutation in the Ca-sensing receptor gene causes familial autosomal dominant hypoparathyroidism. *Am. J. Hum. Genet.* **55(S1):**79, 1994.
80. Guru, S. C., Manickam, P., Crabtree, J. S., et al. Identification and characterization of the multiple endocrine neoplasia type 1 (MEN1) gene. *J. Intern. Med.* **243:**433–439, 1998.
81. Brandi, M. L., Gagel, R. F., Angeli, A., et al. Consensus: guidelines for diagnosis and therapy of MEN type 1 and type 2. *J. Clin. Endocrinol. Metab.* **86:**5658–5671, 2001.
82. Sanso, G. E., Domene, H. M., Rudaz, M. C. G., et al. Very early detection of RET proto-oncogene mutation is crucial for preventive thyroidectomy in multiple endocrine neoplasia type 2 children. *Cancer* **94:**323–330, 2002.
83. Takahasi, M., Asai, N., Iwashita, T., et al. Molecular mechanisms of development of multiple endocrine neoplasia 2 by RET mutations. *J. Intern. Med.* **243(6):**509–513, 1998.
84. Mulligan, L. M., Marsh, D. J., Robinson, B. G., et al. Genotype-phenotype correlation in multiple endocrine neoplasia type 2: report of the International RET Mutation Consortium. *J. Intern. Med.* **238:**343–346, 1995.
85. Carlson, K. M., Dou, S., Chi, D., et al. Single missense mutation in the tyrosine kinase catalytic domain of the RET proto-oncogene is associated with multiple endocrine neoplasia type 2B. *Proc. Natl. Acad. Sci. USA* **91:**1579–1583, 1994.
86. Eng, C., Smith, D. P., Mulligan, L. M., et al. Point mutation within the tyrosine kinase domain of the RET proto-oncogene in multiple endocrine neoplasia type 2B and related sporadic tumours. *Hum. Mol. Genet.* **3:**237–241, 1994.
87. Gimm, O., Marsh, D. J., Andrew, S. D., et al. Germline dinucleotide mutation in codon 883 of the RET proto-oncogene in multiple endocrine neoplasia type 2B without codon 918 mutation. *J. Clin. Endocrinol. Metab.* **82:**3902–3904, 1997.
88. Smith, D. P., Houghton, C., Ponder, B. A., et al. Germline mutation of RET codon 883 in two cases of de novo MEN 2B. *Oncogene* **15:**1213–1217, 1997.
89. Eng, C., Smith, D. P., Mulligan, L. M., et al. A novel point mutation in the tyrosine kinase domain of the RET proto-oncogene in sporadic medullary thyroid carcinoma and in a family with FMTC. *Oncogene* **10:**509–513, 1995.
90. Bolino, A., Shuffenecker, I., Luo, Y., et al. RET mutations in exons 13 and 14 of FMTC patients. *Oncogene* **10:**2415–2419, 1995.
91. Hofstra, R. M., Fattoruso, O., Quadro, L., et al. A novel point mutation in the intracellular domain of the ret proto-oncogene in a family with medullary thyroid carcinoma. *J. Clin. Endocrinol. Metab.***82:**4176–4178, 1997.
92. Berndt, I., Reuter, M., Saller, B., et al. A new hot spot for mutations in the ret proto-oncogene causing familial medullary thyroid carcinoma and multiple endocrine neoplasia type 2A. *J. Clin. Endocrinol. Metab* **83:**770–774, 1998.
93. Rey, J. M., Brouillet, J. P., Fonteneau-Allaire, J., et al. Novel germline RET mutation segregating with papillary thyroid carcinomas. *Genes Chromosomes Cancer* **32:**390–391, 2001.
94. Hoppner, W., Ritter, M. M. A duplication of 12 bp in the critical cysteine rich domain of the RET proto-oncogene results in a distinct phenotype of multiple endocrine neoplasia type 2A. *Hum. Mol. Genet.***6:** 587–590, 1997.
95. Hoppner, W., Dralle, H., and Brabant. G. Duplication of 9 base pairs in the critical cysteine-rich domain of the RET proto-oncogene causes multiple endocrine neoplasia type 2A. *Hum. Mutat.* **1 (Suppl. 1):**S128–S130, 1998.
96. Lips, C. J., Landsvater, R. M., Hoppnener, J. W., et al. Clinical screening as compared with DNA analysis in families with multiple endocrine neoplasia type 2A. *N. Engl. J. Med.* **331:**828–835, 1994.
97. Bartsch, D. K., Hasse, C., Schug, C., et al. A RET double mutation in the germline of a kindred with FMTC. *Exp. Clin. Endocrinol. Diabetes* **108:**128–132, 2000.
98. Miyauchi, A., Futami, H., Hai, N., et al. Two germline missense mutations at codons 804 and 806 of the RET proto-oncogene in the same allele in a patient with multiple endocrine neoplasia type 2B without codon 918 mutation. *Jpn. J. Cancer Res.* **90:**1–5, 1999.
99. American Academy of Pediatrics Committee on Bioethics. Policy Statement: Ethical issues with genetic testing in pediatrics. *Pediatrics* **107:**1451–1455, 2001.
100. ASHG/ACMG. Points to consider: ethical, legal, and psychological implications of genetic testing in children and adolescents. *Am. J. Hum. Genet.* **57:**1233–1241, 1995.
101. The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care* **22(S1):**S5–S19, 1999.
102. Stride, A., and Hattersley, A. T. Different genes, different diabetes: lessons from maturity-onset diabetes of the young. *Ann. Med.* **34:**207–216, 2002.
103. Fajans, S. S., Bell, G. I., and Polonsky, K. S. Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. *N. Engl. J. Med.* **345:**971–980, 2001.
104. Velho, G., and Robert, J. J. Maturity-onset diabetes of the young (MODY): genetic and clinical characteristics. *Horm. Res.* **57:**29–33, 2002.

105. Frayling, T. M., Evans, J. C., Bulman, M. P., et al. β-Cell genes and diabetes: molecular and clinical characterization of mutations in transcription factors. *Diabetes* **50(Suppl. 1):**S94–S100, 2001.
106. Matschinsky, F. M., Glaser, B., and Magnuson, M. A. Pancreatic β-cell glucokinase: closing the gap between theoretical concepts and experimental realities. *Diabetes* **47:**307–315, 1998.
107. Froguel, P., Vaxillaire, M., Sun, F., et al. Close linkage of glucokinase locus on chromosome 7p to early-onset non-insulin-dependent diabetes mellitus. *Nature* **356:**162–164, 1992.
108. Vionnet, N., Stoffel, M., Takeda, J., et al. Nonsense mutation in the glucokinase gene causes early-onset non-insulin-dependent diabetes mellitus. *Nature* **356:**721–722, 1992.
109. Yamagata, K., Furuta, H., Oda, N., et al. Mutations in the hepatocyte nuclear factor-4alpha gene in maturity-onset diabetes of the young (MODY1). *Nature* **384:**458–460, 1996.
110. Ryffel, G. U. Mutations in the human genes encoding the transcription factors of the hepatocyte nuclear factor (HNF)1 and HNF4 families: functional and pathological consequences. *J. Mol. Endocrinol.* **27:**11–29, 2001.
111. Thomas, H., Jaschkowitz, K., Bulman, M., et al. A distant upstream promoter of the HNF-4α gene connects the transcription factors involved in maturity-onset diabetes of the young. *Hum. Mol. Genet.* **10:**2089–2097, 2001.
112. Gloyn, A. L., Ellard, S., Shepherd, M., et al. Maturity-onset diabetes of the young caused by a balanced translocation where the 20q12 break point results in disruption upstream of the coding region of hepatocyte nuclear factor-4α (HNF4A) gene. *Diabetes* **51:** 2329–2333, 2002.
113. Yamagata, K., Oda, N., Kaisaki, P. J., et al. Mutations in the hepatocyte nuclear factor-1-alpha gene in maturity-onset diabetes of the young (MODY3). *Nature* **384:**455–458, 1996.
114. Kaisaki, P. J, Menzel, S., Lindner, T., et al. Mutations in the hepatocyte nuclear factor-1 alpha gene in MODY and early-onset NIDDM: evidence for a mutational hotspot in exon 4. *Diabetes*; **46:** 528–535, 1997.
115. Stoffers,D. A., Ferrer, J., Clarke, W. L., et al. Early-onset type II diabetes mellitus (MODY4) linked to IPF1. *Nat. Genet.* **17:**138–139, 1997.
116. Hui, H., and Perfetti, R. Pancreas duodenum homeobox-1 regulates pancreas development during embryogenesis and islet cell function in adulthood. *Eur. J. Endocrinol.* **146:**129–141, 2002.
117. Stoffers, D. A., Zinkin, N. T., Stanojevic, V., et al. Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence. *Nat, Genet.* **15:**106–110, 1997.
118. Horikawa, Y., Iwasaki, N., Hara, M., et al. Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY. *Nat. Genet.* **17:**384–385, 1997.
119. Nishigori, H., Yamada, S.,Kohama, T., et al. Frameshift mutation, A263fsinsGG, in the hepatocyte nuclear factor-1β gene associated with diabetes and renal dysfunction. *Diabetes* **47:**1354–1355, 1998.
120. Lindner, T. H., Njølstad, P. R, Horikawa, Y., et al. A novel syndrome of diabetes mellitus, renal dysfunction and genital malformation associated with a partial deletion of the pseudo-POU domain of hepatocyte nuclear factor-1β. *Hum. Mol. Genet.* **8:**2001–2008.
121 .Malecki, M. T., Jhala, U. S., Antonellis, A., et al. Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus. *Nat. Genet.* **23:**323–328, 1999.
122. Naya, F. J., Huang, H. P., Qiu, Y., et al. Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/NeuroD-deficient mice. *Genes Dev.***11:** 2323–2334,1997.
123. Delépine, M., Nicolino, M., Barrett, T., et al. EIF2AK3, encoding translation initiation factor 2-α kinase 3, is mutated in patients with Wolcott–Rallison syndrome. *Nat. Genet.* **25:**406–409,2000.
124. Shi, Y., Taylor, S. I., Tan, S. L., et al. When translation meets metabolism: multiple links to diabetes. *Endocr. Rev.* **24:**91–101,2003.
125. Strom, T. M., Hörtnagel, K., Hofmann, S., et al. Diabetes insipidus, diabetes mellitus, optic atrophy and deafness (DIDMOAD) caused by mutations in a novel gene (*wolframin*) coding for a predicted transmembrane protein. *Hum. Mol. Genet.* **7:**2021–2028, 1998.
126. Inoue, H., Tanizawa, Y., Wasson, J., et al. A gene encoding a transmembrane protein is mutated in patients with diabetes mellitus and optic atrophy (Wolfram syndrome). *Nat. Genet.* **20:**143–148, 1998.
127. Khanim, F., Kirk, J., Latif, F., et al. WFS1/Wolframin mutations, Wolfram syndrome, and associated diseases. *Hum. Mutat.* **17:**357–367, 2001.
128. Van den Ouweland, J. M., Lemkes, H. H., Ruitenbeek, W., et al. Mutation in mitochondrial tRNA(Leu)(UUR) gene in a large pedigree with maternally inherited diabetes mellitus and deafness. *Nat. Genet.* **1:**368-371, 1998.
129. Velho, G., Byrne, M. M., Clement, K., et al. Clinical phenotypes, insulin secretion, and insulin sensitivity in kindreds with maternally inherited diabetes and deafness due to mitochondrial tRNA sup Leu(UUR) gene mutation. *Diabetes* **45:**478–487, 1996.
130. Redondo, M. J., and Eisenbarth, G. S. Genetic control of autoimmunity in type 1 diabetes and associate disorders. *Diabetologia* **45:**605–622, 2002.
131. The Finnish-German APECED Consortium. An autoimmune disease, APECED, caused by mutations in a novel gene featuring two PHD-type zinc-finger domains. *Nat. Genet.* **17:**399–403, 1997.
132. Nagamine, K., Peterson, P., Scott, H. S., et al. Positional cloning of the APECED gene. *Nat. Genet.* **17:**393–398, 1997.
133. Chatila, T. A., Blaeser, F., Ho, N., et al. JM2, encoding a fork head-related protein, is mutated in X-linked autoimmunity-allergic disregulation (sic) syndrome. *J. Clin. Invest.* **106:**R75–R81, 2000.
134. McCarthy, M., and Menzel, S. The genetics of type 2 diabetes. *Br. J. Clin. Pharmacol.* **51:**195–199, 2001.
135. Medici, F., Hawa, M., Ianari, A., et al. Concordance rate for type II diabetes mellitus in monozygotic twins: actuarial analysis. *Diabetologia* **42:**146–150, 1999.
136. Van Tilburg, J., van Haeften, T. W., Pearson, P., et al. Defining the genetic contribution of type 2 diabetes mellitus. *J. Med. Genet.* **38:** 569–578, 2001.
137. Horikawa, Y., Oda, N., Cox, N. J, et al. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat. Genet.* **26:**163–175, 2000.
138. Cox, N. J. Challenges in identifying genetic variation affecting susceptibility to type 2 diabetes: examples from studies of the calpain-10 gene. *Hum. Mol. Genet.* **10:**2301–2305, 2001.
139. Hirschhorn, J. N., Lohmueller, K., Byrne, E., et al. A comprehensive review of genetic association studies. *Genet. Med.* **4:**45–61, 2002.
140. Lohmueller, K. E., Pearce, C. L., Pike, M., et al. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat. Genet.* **33:**177–182, 2003.

25 Molecular Pathogenesis of Cardiovascular Disease

ALAN H. B. WU

1. INTRODUCTION

This chapter summarizes the large volume of relatively recent work devoted to finding genetic polymorphisms that are linked to the development of cardiovascular disease (CVD) and acute coronary syndromes (ACS). These markers might become part of a panel of laboratory tests and clinical indicators that will be used for predicting future CVD risk. The medical literature is exploding in the last 10 yr with genotype–disease association studies. As such, the criteria for establishing high-quality genetic association studies are now needed and have been opined by some. One such list is shown in Table 1 *(1)*. With specific regard to polymorphic studies with CVD, Jones and Montgomery described necessary additional steps *(2)* (Table 2). Candidate polymorphisms should be directed against genes that produce products known to be involved in the pathophysiology of cardiac diseases. In the case of ACS, this would include (but not be limited to) genes that regulate inflammation, thrombosis, platelet function, lipoproteins, and other markers of atherosclerosis. In the case of congestive heart failure (CHF), this would include genes that regulate the renin–angiotensin–aldosterone axis, natriuretic peptides, and adrenoreceptors. Having identified polymorphisms in candidate genes, it would be desirable to show that the variants produce different concentrations and/or functional capabilities of the expressed protein. This can demonstrated by an in vitro cell culture model or in human studies in which the concentration of the protein in the variant is compared against that found in the wild type. Having fulfilled one or both of these criteria, a clinical study can be conducted that compares the incidence the the polymorphisms in diseased and carefully selected control populations. The phenotype of the diseased population should be tightly controlled. It is not appropriate to combine ischemic cardiomyopathy patients with CHF patients who have a valvular etiology. It is also important to identify and exclude confounding variables such as the use of cardiac medications that modify the expression and concentration of the target protein. As most studies to date have only shown a modest relative risk ratio, the number of subjects enrolled should be sufficiently powered to detect a small but significant difference. Meta-analyses are important to increase the enrollments necessary to achieve the discriminatory power.

Table 3 lists some of the major single-nucleotide polymorphisms (SNPS), variable number of tandem repeats (VNTRs) and insertion/deletion mutations genes that are discussed in this chapter. The identification of specific and relevant genotypes might ultimately have an impact beyond risk stratification. The future emphasis of polymorphisms in cardiac disease will be focused on pharmacogenomics (i.e., the determination of genotypes prior to the selection of drugs for treatment). In the future, it might be necessary to document a certain genotype for the proper selection of statins for primary CVDs, drugs for hypertension, angiotensin-converting enzyme and β-blockers for heart failure, or glycoprotein IIb/IIIa inhibitors in ACS. Although this field is in its infancy, CVDs might be one area that would benefit greatly from this technology.

2. GENES OF THROMBOTIC OR THROMBOLYTIC FUNCTION

2.1. FACTORS V AND II The backround on hemostasis, molecular mechanisms for venous thrombosis, and description of the common mutations for Factors V and II are given in Chapter 41. Given the important role of Factor V Leiden in venous thrombosis, studies were conducted to determine if this polymorphism is a risk factor for arterial thrombosis. Surprisingly, the data have consistently shown that the presence of Factor V Leiden is not associated with an increased incidence of coronary artery disease (CAD). In the Physician Health Study, the odds ratio was not statistically significant for development of acute myocardial infarction (AMI) when all subjects were examined, or within various subgroups (i.e., age ≤60 yr, nonsmokers, and absence of familial history, hypercholesterolemia, or hypertension) *(3)*. A meta-analysis of 5431 CAD cases and controls from 9 published studies confirmed these data (odds ratio: [OR] 1.24; 95% confidence interval [CI]: 0.84–1.59) *(4)*. In terms of venous thrombosis, the G20210A polymorphism is not associated with an increased

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

Table 1
Criteria for Genetic Association Studies

Well-defined phenotype
A clear *a priori* hypothesis
Large sample size and *a priori* power analysis
Small *p*-value ($p < 0.05$) and high odds ratio
Biological plausibility of the hypothesis
Independent replication of results in different populations
Confirmation in family-based studies with at least two other common polymorphisms within the same gene

Source: Adapted from ref. *1*.

Table 2
Template for Studies in the Correlation of Polymorphism and CDV Risk

The polymorphic variant should have in vitro functionality.
The polymorphic variant should have in vivo functionality.
The phenotype should be very tightly defined.
The use of confounding agents should be excluded.
The study should be appropriately powered.

Source: Adapted from Jones, A. and Montgomery, H. *Eur. Heart J.* **23:**1071–1074, 2002.

incidence of ACS. In a meta-analysis of 5607 CAD cases and controls, the odds ratio was 1.15 (95% CI: 0.84–1.59) *(5)*.

2.2. FACTOR VII The rupture of a coronary artery results in the exposure of tissue factor to Factor VII in blood that becomes activated and initiates the extrinsic coagulation pathway. High plasma concentrations of Factor VII have been implicated as a risk factor to CAD *(5)*. There are several common polymorphisms in the gene for Factor VII. One of the more widely studied is the R353Q involving a substitution of glutamine for arginine. The allele frequency is about 13%, with a homozygous rate of about 2%. The allele frequency among Asians is lower at about 6%. Girelli et al. found that subjects with the AA wild-type genotype had a higher Factor VIIa concentration (50.9 mU/mL) than subjects with the heterozygous AG (31.5 mU/mL) and the homozygous GG (14.0 mU/mL) genotypes *(6)*. The lower Factor VIIa concentration in subjects with the G allele would suggest that this mutation would offer a protective effect against ACS. This has been demonstrated in several reports correlating the presence of Factor VII genotypes with incidence of ACS. Although many of the initial studies did not have the number of subjects to demonstrate statistical significance, a meta-analysis of 2574 CAD cases and controls produced an odds ratio of 0.78 (95% CI: 0.65–0.93) for the AG and GG genotypes *(3)*. Yet undeveloped therapies might be directed at lowering Factor VII concentrations in individuals with the AA genotype.

There are other mutations in Factor VII that have been identified. In intron 7, there is a size polymorphism resulting from the presence of a 37-bp repeat. The wild-type sequence contains 6 copies of the 37-base repeat sequence. A less common allele is the presence of seven repeats, whereas the rare alleles contain five and eight repeats. In one study, the allele frequencies were 66%, 33%, 0.7%, and <0.1% for the five to eight repeats, respectively *(6)*. There is also an insertion polymorphism in the 5′ promoter region of the Factor VII gene, which corresponds to the addition of 10 bp. The allele frequency for the insertion mutation is about 18%. Although the significance toward CAD for these polymorphisms have not been thoroughly studied as yet, preliminary data do not suggest that there will be a major correlation with arterial thrombosis and determination of Factor VII plasma concentrations.

2.3. FIBRINOGEN Increased plasma fibrinogen concentrations are linked to CAD. In the Northwick Park Heart Study, an increase of 60 mg/dL confers an 84% increase in the risk of CAD over 5 yr *(5)*. Smoking also increases the fibrinogen concentration and is an established environmental risk factor. Plasma fibrinogen concentrations and cardiovascular risk might also be linked to genetic factors. There are at least eight fibrinogen gene polymorphism that have recently been described in the α-, β-, and γ-chains *(7)*. The one most commonly studied is in the B-fibrinogen promoter region at position –455, where there is a change from a G→A. Production of the B-chain is the rate-limiting step in the formation of mature fibrinogen. The A allele frequency is 20%. The 455 A genotype is associated with an increased plasma fibrinogen concentration (370 mg/dL) vs the wild type (G455, 320 mg/dL), as shown in one study *(8)*. However, other studies have rather consistently shown that there is no association of the fibrinogen polymorphism and presence of or risk for CAD *(9)*.

Genotyping for the –455 polymorphism might be useful in selecting patients who might benefit from statin therapy. de Maat et al. showed that CAD patients with the –455 AA genotype treated with pravastatin had less disease progression (as measured by coronary artery mean segment and minimum obstruction diameters by angiography) than the –455AG and –455GG genotypes *(10)*. Whether this evolves into some routine pharmacogenomic practice in CAD patients or hyperlipidemic subjects remains to be seen.

2.4. PLASMINOGEN ACTIVATOR INHIBITOR-1 Interest in polymorphisms in plasminogen activator inhibitor-1 (PAI-1) stems in part from the observation that there is reduced fibrinolytic function and increased risk for CAD in the presence of increased blood concentrations of PAI-1, because of an accelerated inactivation of tissue plasminogen activator (tPA), a natural agent for thrombolysis. Paradoxically, increased plasma tPA concentrations are also associated with risk for myocardial infarction *(11)*. In the European Concerted Action on Thrombosis and Disabilities Angina Pectoris Study, the blood of 3043 patients with angina were measured for tPA and PAI-1. Patients who subsequently suffered MI or sudden coronary death had higher concentrations of PAI-1 and tPA (18.2 and 11.8 ng/mL, respectively) than those who did not suffer an event (14.8 and 10.0 ng/mL, respectively) *(12)*. It is possible that although tPA has a fibrinolytic role, it might also destabilize the atherosclerotic plaque.

An insertion/deletion polymorphism is has been identified in the promoter region of the PAI-1 gene whereby one allele sequence has four guanosines (4G) and the other has five (5G). Both alleles bind a transcriptional activator. However, the 5G allele also binds a repressor protein to an overlapping binding site, thereby reducing the level of transcription. As a consequence, the 4G allele has been related to higher PAI-1

Table 3
Summary of Single-Nucleotide Polymorphism (SNP) and Representative Restriction Endonuclease Used in Detection

Factor	*Gene*	*SNP location*	*Endonuclease*[a]
II	3′ untranslated region,	20210 G→A, arg→gln	*Hin*dIII
V	Leiden gene, exon 10	1691 G→A, arg→gln	*Mnl*I
VII	Catalytic region, exon 8	353 G→A, arg→gln	*Msp*I
	Hypervariable region, intron 7	VTNR, 37 bp	*Rsa*I
	5′F7, 5′ promoter region	10-bp insertion	NA
Fibrinogen	β-Chain promoter region	–455 G→A	*Hae*III
PAI-1	Promoter region	–675, G insertion/deletion	NA
GP Ia	Coding region, intron 7	807 C→T, no sub.[b]	NA
	Coding region, intron 7	873 G→A, no sub.	NA
	Cation binding domain	1648G→A, glu→lys	*Mnl*I
GP Ibα	Hpa-2, Ko polymorphism	434 C→T, thr→met	*Bsa*HI
	HPA-2	VTNR, 39 bp	
	Initiation codon, HPA-2, Kozak	–5 from ATG start, T→C no sub.[b]	
GP IIb	HPA-3, heavy chain	843 T→G, ile→ser	*Bse*GI or *Fok*I
GP IIIa	HPA-1, exon 2	1565 C→T, leu→pro	*Nci*I
GP VI	Exon 5	13254 T→C, ser→pro	*Hpa*II
MTHFR	Encoding region	677 C→T, ala→val	*Hinf*I
Lipoprotein	Intron 8	495 T→G	*Hin*dIII
Lipase	Intron 6	C→T	*Pvu*II
	Exon 9	447 C→G, ser→term	*Mnl*I
	Exon 2	9, G→A, asp→asn	*Bcl*I
	Exon 6	291 A→G asn→ser	*Rsa*I
Paraoxonase	PON1	55 leu→met	*Hsp*92II
	PON1	192 gln→arg	*ALw*I
	PON3	311 cys→ser	*Dde*I
NADH oxid.	$p22^{phox}$	242 C→T, tyr→his	*Rsa*I
NO synthase	Exon 7 of NOS3	894 G→T, glu→asp	*MBo*I
	Intron 4 of NOS3	VNTR, 27 bp	NA
	5′ Promoter region of NOS3	–786 T→C	*Hae*III
Apo E	ε2	118, 158 cys	*Hha*I
	ε3	118 cyst, 158 arg	
	ε4	118, 158 arg	
ACE	Intron 16	287-bp insertion/deletion	NA
Angiotensin receptor	5′ end of the 3′ untranslated region	1166 A→C	*Dde*I
Adrenergic receptor			
α_2	Intracellular domain of α_2C	322–325 4-bp deletion	*Nci*I
β_1	N-terminal of coding exon	145 A→G Ser49Gly	*Eco*0109I
	Transmembrane helix VII region	1165 G→C Arg389Gly	*Bcg*I
β_2	fourth transmembrane-spanning domain	Arg16Gly	*Sty*1 or *Eco*130I
		Gln27Glu	*Fnu*4H1
		Thr164Ile	*Mnl*1

[a]Representative restriction endonuclease. Other appropriate ones could be used.
[b]No sub.= Polymorphism does not result in an alteration in gene expression; NA= not applicable.

concentrations in plasma *(13)*. The 5G allele frequency is about 48% among Caucasians and is slightly lower among Asians.

Studies on the role of PAI-1 polymorphism in ACS have been equivocal. Most studies, however, had low numbers (<1000) of enrolled subjects. In a meta-analysis, Iacoviello et al. combined results of 9 studies on 3641 CAD cases and controls. A positive association between the 4G/4G allele for CAD vs controls was obtained (OR: 1.30; 95% CI: 1.07–1.58) *(14)*. For most of the individual studies that were included, the odds ratio did not produce a statistically significant association for the 4G allele and CAD (i.e., a positive association observed only when studies were combined). For genetic risk assessment markers that might only have a borderline association, enrollment bias such as history of CAD might play a major role in the accuracy of results. For example, in an age-matched control population, how does the individual or investigator verify that silent ischemic disease had not occurred? As such, Ossei-Gerning et al. used coronary angiography as a criterion for CAD and found a positive association of the 4G genotype with the presence or absence of multivessel disease *(15)*.

3. PLATELET GLYCOPROTEINS

Platelets work in concert with coagulation factors to maintain hemostasis. In ACS, platelet activation plays a major role in the arterial thrombosis of coronary arteries. Whereas a fibrin clot dominates in patients with totally occluded ST-segment AMI, platelet clots are more responsible for the partial occlusions observed in non-ST-segment AMI and unstable angina *(16)*. There are several glycoprotein (GP) receptor complexes found on the platelet surface that facilitate the binding functions of

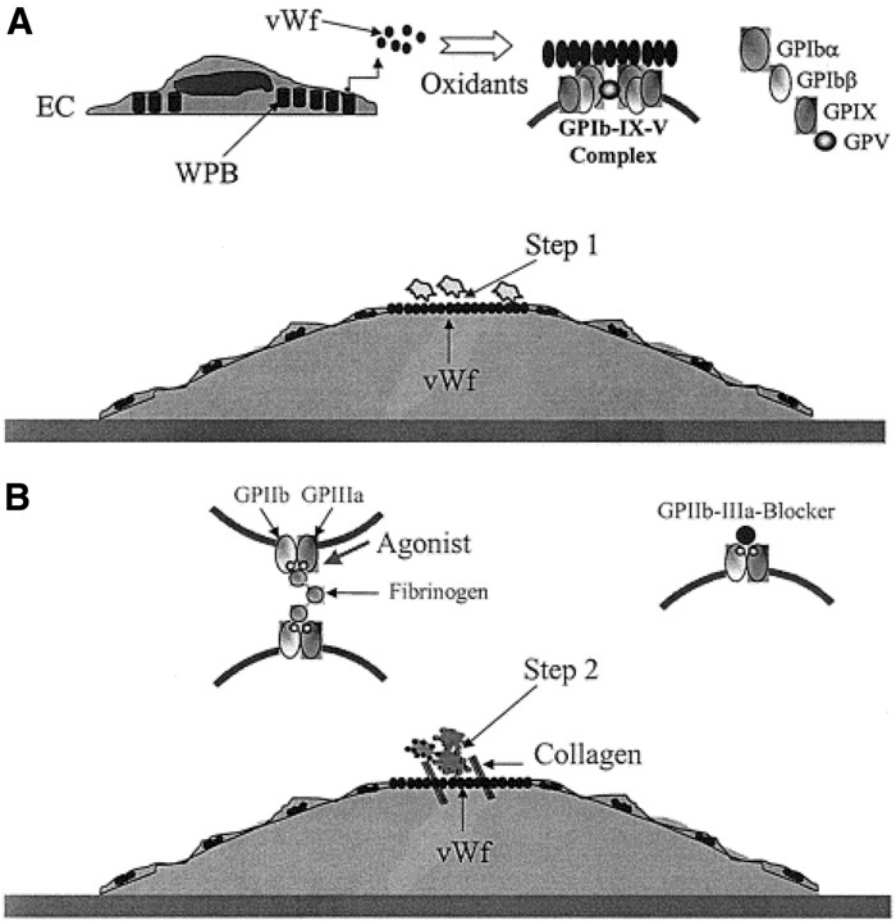


Fig. 1. (A) The initial phase of primary heomstasis. Platelets bind to the GP Ib-IX-V receptor complex to von Willebrand factor (vWf) under high stress on the subendothelial matrix. Weibel–Palade (WPB) bodies from the endothelial cells bind to the GPIb-IX-V receptor. **(B)** The secondary phase of hemostasis. In the presence of collagen, platelets are activated causing expression of the GP IIb/IIIa receptor, which stimulates platelets to aggregate. (Used with permission from Kandzari, D. E., and Goldschmidt-Clermont, P. J. Platelet polymorphisms and ischemic heart disease: moving beyond traditional risk factors. *J. Am. Coll. Cardiol.* **38:**1028–1031, 2001.)

platelets *(17).* The role of these glycoproteins is shown in Fig. 1. GP Ib-V-IX receptor binds to the von Willebrand factor receptor and is involved with the adhesion of platelets to the subendothelium, particularly under the conditions of high sheer stress. GP Ia-IIa receptor binds to collagen and also stimulates platelet adhesion to endothelial surfaces. GP IIb/IIIa is the most abundant receptor and functions to bind fibrinogen. Sheer forces from systole and many different agonists can activate the GP IIb/IIIa receptor. Once activated, aggregation occurs by the binding of multiple platelets to the fibrinogen molecule. The GP IIb/IIIa receptor is of particular interest for cardiologists, as inhibitors to the receptor are used to treat patients with ACS and as a prophylactic measure for patients undergoing percutaneous coronary intervention (PCI).

Specific mutations to either GP IIb or IIIa might result in a reduced or possibly enhanced binding to of platelets to fibrinogen. A widely studied polymorphism is in the GPIIIa receptor, a change from cytosine to thymidine at nucleotide 1565, resulting in the substitution from a leucine to a proline. The A2 allele frequency is 16% with a 2% incidence of homozygotes among Caucasians. The allele frequency at 7% is lower among Asians. This particular mutation has been implicated in immune-mediate platelet destruction, an allogen referred to as PI^{A2}. As with other polymorphisms studied for arterial thrombosis, the literature contains contradictory findings *(18,19).* However, for the GP IIIa polymorphism, a meta-analysis of 9 studies on 7920 CAD cases and controls produced an odds ratio of 1.12, which barely reached statistical significance (95% CI: 1.01–1.24) *(3).*

A recent study has suggested that the presence of the PI^{A2} polymorphism might correlate to a higher degree of myocardial injury after elective cardiopulmonary bypass surgery *(20).* Using cardiac troponin I as a marker of myocardial damage, patients with one or two PI^{A2} alleles had more cardiac injury than the wild type. The authors suggested that the patients with the polymorphism had enhanced platelet adhesion and aggregability, creating a more robust platelet plug. This data might have an impact on patients who underwent emergency bypass surgery after a failed coronary angioplasty procedure. Although treatment with glycoprotein IIb/IIIa inhibitors is not warranted during cardiac bypass because of the potential for bleeding *(21),* it is widely used prior to angioplasty. Patients with the PI^{A2} genotype might benefit from prophylactic platelet transfusions *(22).*

Given the importance of GP IIb/IIIa receptor inhibitors as therapeutic agents in patients with ACS, it might be important to determine if GP IIb/IIIa polymorphisms affect the success of these drugs. One study clearly showed that the PI^{A2} genotype following abciximab use was associated by reduced platelet inhibition and suggested that it might contribute to an unfavorable outcome *(23)*. Studies that attempted to correlate GP IIIa polymorphisms with clinical outcomes have been published. Laule et al. *(24)* found no association of the A2 allele for the 30-d composite end points (target–vessel revascularization, AMI, and death) when 653 cases were examined. In contrast, Kastrati et al.reported higher restenosis rates at 6 mo for the A2 allele (OR: 1.35; 95% CI: 1.07–1.70) among 1150 coronary angioplasty patients *(25)*. Unfortunately, the effect of GP IIIa polymorphisms on the success of GP IIb/IIIa inhibitor therapy were not examined, as these inhibitors were not given to patients in either of these studies. Given the low rate of adverse events for patients given GP IIb/IIIa inhibitors, a fairly large clinical trial will be necessary to document the effect. Such a study is unlikely to be funded by a pharmaceutical company, as it might view such testing as a hinderance for widespread use of these drugs.

There has been published data on the polymorphisms of other platelet glycoproteins. GP IIb is designated as the human platelet alloantigen(HPA)-3. A polymorphism exists at position 943, where there is a change from T→G, resulting in the substitution of an isoleucine for a serine. The allele frequency for the mutation is about 40% with 12–15% homozygotes. Preliminary studies suggest that the HPA-3 polymorphism is not associated with the presence of CAD *(26)*.

Glycoprotein Ia (also known as integrin $\alpha_2\beta_1$) is part of the GP Ia/IIa heterodimer and functions to adhere platelets to injured vessel walls, and it is designated as the human platelet allogen-5. There are two silent polymorphisms present in the GP Ia gene. At position 807, there is a C→T substitution, and at 873, there is a G→A substitution. These polymorphisms are linked. In both cases, there is no change in the amino acid sequence. Although silent mutations do not result in the alteration of the actual glycoprotein product, they might be associated with mutations in regulatory regions, such as the gene promoter. Alternately, changes in the DNA sequence might influence the survival of mRNA such that the amount of protein produced is altered. The detection of these polymorphisms is achieved by use of allele-specific fluorescence probes. Other polymorphisms have been identified at nucleotide 1648, where there is a G→A change that results in the substitution of glutamine for lysine. This polymorphism is linked to the C807T gene, with the AA1648 found exclusively with CC807 and GG1648 with TT807. There is a rare polymorphism (frequency <0.01) at nucleotide 2531 involving a C→T substitution.

Furihata et al. identified three major alleles (and one rare allele) combining polymorphisms a nucleotides 807, 1648, and 2531 *(27)*. Using flow cytometry and specific monoclonal antibodies, these investigators showed that these alleles expressed different densities of receptors (summarized in Table 4). The frequencies of these alleles are 39%, 53%, and 7.6%. Heterozygote mixtures of these alleles produce receptor densities that are intermediate to the homozygote genotypes *(28)*. The presence of a high density of receptors (allele 1) might correlate to a greater degree of platelet adhesion onto collagen and the endothelial wall, thereby stimulating a hyperthrombotic response.

Table 4
Alleles of the GPIa Gene and Correlation to Receptor Levels

Allele No.	*Genotype*	*Receptor density*
1	807T/1648G/2531C	High
2	807C/1648G/2531C	Intermediate
3	807C/1648A/2531 C	Low

Source: Data from ref. *27*

Studies have been conducted to determine if GP Ia polymophorism is associated with an increased incidence of ACS or poor outcomes in patients treated with PCI. In a study of young patients (<49 yr) who survived myocardial infarction, there was a significant association of the T807 allele of GP Ia *(29)*. It was hypothesized that this polymorphism is more effective as a risk factor among young patients, which might get "diluted" with the presence of other risk factors (e.g., as the age of the patient increases). In contrast, a study of Finnish men with sudden cardiac death and AMI showed no association of this genotype *(30)*. GP Ia polymorphism was studied as a risk factor for patients undergoing angioplasty. Restenosis and the need for revascularization can be caused by stent placement, as it can activate platelets to form thrombosis and promote the migration and growth of smooth muscle cells. However, von Beckerath et al. found no significant difference in the 30-d incidence of death, AMI, or urgent target–vessel revascularization and polymorphisms among 1797 patients undergoing PCI *(31)*. The authors postulate that the mechanism of platelet activation following PCI might explain these findings. Stents are usually placed in high-grade lesions that contain more collagen-rich plaques as opposed to low-grade lesions, which are richer in platelet tissue factor. If the restenosis and other adverse events occurred in these low-grade areas, GP Ia polymorphism might not be a factor.

In contrast, a study of 2163 Caucasian males showed that the presence of the A1648 allele of GPIa was associated with an increased incidence of CAD when the analysis was restricted to low-risk subgroups. This association was independent to the C807T genotype. There was no statistical difference in genotype frequency and incidence of CAD when all patients were studied. It might be possible that the role of platelet polymorphism could be minimized or superceded by the presence of established CAD risk factors.

Polymorphisms also exist in the third of the three platelet receptors (i.e., GP-Ib-IX-V). GP Ib is associated with the HPA-2 alloantigen. The Ko polymorphism is a single-nucleotide polmorphism(SNP) at nucleotide 434 involving a C→T change producing a substitution of a threonine to a methionine at codon 145. The allele frequency for the 434T is 12% *(32)*. A VTNR polymorphism of 39 bp is present in GP Ibα, resulting in 13 amino acids that is repeated between one and four times *(33)*. Both of these polymorphisms have the potential to affect the structure of GPIbα and appear to have strong linkage disequilibrium. The allele frequency is 8.3%, 78%, 13%, and 1%

for the one to four repeats, respectively (labeled D, C, B, and A). The hypothesis is that an increased number of repeats predisposes the platelet receptor to binding to endothelial surfaces and thrombosis. The Kozak sequence polymorphism results in the change of a T→C at position –5 from the ATG start codon. Although this does not alter the mature glycoprotein, it might influence the extent of the gene expression. The frequency of the –5C allele is 15%.

As with the other studies on platelet polymorphisms, the incidence of CAD and complications following angioplasty and presence of GP Ibα genotypes is equivocal. Mikkelsson et al. found that the 145M genotype of HPA-2 was associated with the VNTR B (three variable repeats) haplotype, and this combination was a significant risk factor for coronary thrombosis among Finnish men who suffered sudden death AMI or coronary thrombosis or died of noncardiac causes *(34).* The odds ratio was highest among young men (OR: 9.2; 95% CI: 2.4–35.0). Meisel et al. found no difference in the incidence of the C-5 allele between CAD and control groups *(35).* However, when 269 patients undergoing PCI were examined, the presence of the –5C allele was associated with a higher incidence of 30-d adverse events. Interestingly, there was no difference among the 103 patients who underwent directional athrectomy and 278 in whom a stent was placed. It is unclear why the PTCA group differed, considering all of the procedures have risks for thrombosis and restenosis. For the Kozak sequence polymorphism, Corral showed no correlation as a risk factor for arterial thrombosis among 101 CAD patients *(36).*

A promising new polymorphic marker is glycoprotein VI, which is important in the activation of phospholipase Cγ2. A change in nucleotide 13254 of a T→C results in the substitution of a serine to a proline. The allele frequency of the 13254C allele is 16%. In a study of AMI patients, the overall odds ratio for the presence of the CC genotype was not significant vs controls. However, in various subgroups, the odds ratio was 4.52 (95% CI: 1.23–16.6) and 6.48 (95% CI: 1.47–28.5) for women and patients ≥60 years *(37).* The mechanism of this interaction remains to be determined.

4. BIOCHEMICAL FACTORS

4.1. LIPOPROTEIN LIPASE Plasma lipid and lipoprotein concentrations are important factors linked to CVD risk. In addition to the well-studied dietary factors controlling plasma lipid concentrations, there is growing evidence that genetics plays a factor. Polymorphisms in lipoprotein lipase (LPL) is one gene that has been studied in this regard. The LPL gene is found on chromosome 8p33 and contains 10 exons spanning 30 kb. This gene produces a 475-amino-acid protein that is posttranslationally modified to a signal peptide and the 448-amino-acid mature enzyme that has a molecular weight of about 60 kDa. LPL is anchored to the vascular endothelium and removes lipids from the circulation by hydrolyzing triglycerides present in chylomicrons and very low-density lipoprotein(VLDL). Familial LPL deficiency causes marked chylomicronemia and premature atherosclerosis, but is relatively rare. There are at least four polymorphic variants of the LPL gene that have a more subtle effect on plasma lipid concentrations and cardiovascular risk. If these polymorphisms are associated with a reduced activity of LPL, they might be genetic risk factors for atherosclerosis.

Two of the most widely studied polymorphisms to lipoprotein lipase are the restriction fragment length polymorphic sites *Hin*dIII and *Pvu*II, located on introns 8 and 6, respectively. The *Hin*dIII site results in a T→G substitution and has an allele frequency (presence of the restriction site, H+) of roughly 70%. The *Pvu*II site results in a C→T substitution and has an allele frequency (P+) of about 45%. Given the importance of lipoprotein lipase to lipid metabolism, attempts have been made to correlate the presence of these polymorphisms to lipid levels. Gerdes et al. showed that subjects with the H^+H^+ genotype had higher triglyceride (mean = 125 mg/dL) and lower high-density lipoprotein (HDL) cholesterol concentrations (mean = 49 mg/dL) than those with the H^-H^- genotype (110 and 53 mg/dL, respectively) *(38).* Larson et al. put enrolled subjects on the same low-fat and low-cholesterol, and high-carbohydrate and high-fiber diets prior to lipid testing, and found no difference in *Hin*dIII genotypes for triglycerides and HDL cholesterol *(39).* The total and LDL cholesterol, however, was increased for women in the H^+H^+ group. Other investigators found similar results for triglycerides for the *Pvu*II polymorphism (158 for P^-P^- vs 213 mg/dL for P^+P^+) *(40).*

The presence of the H^+ allele is associated with a higher frequency of coronary disease. When comparing *Hin*dIII polymorphism to the number of disease coronary arteries, the H^+H^+ genotype had a higher incidence of multivessel disease (OR: 4.4; 95% CI: 1.73–11.33) than the combined H^+H^- and H^-H^- genotypes among Italians patients surviving AMI *(41).* Because the H^+ allele has the higher prevalence, the presence of the H^- variant would be interpreted to offer a protective effect toward CAD. Similar but less dramatic results were obtained for the *Pvu*II polymorphism, where the P^+P^+ genotype had an odds ratio of 1.73 (95% CI: 1.03–2.89) vs the P′P′ genotype *(40).* Studies examining the risk of particular H and P genotypes with CAD are being initiated. In one study, the H^+H^+ genotypes was significantly associated with an odds ratio of 2.0 (95% CI: 1.11–3.70) vs the H′H′ genotype in 725 patients undergoing angioplasty for ACS and 168 control subjects with normal coronary arteries *(42).* This same study produced a trend for the P^+P^+ vs the P^-P^- genotypes, but results did not reach significance (OR: 1.39; 95% CI: 0.84–2.28).

Another widely studied polymorphism in lipoprotein lipase is the point mutation of a C→G in nucleotide 447 of exon 9, resulting in the substitution of serine for a permature termination codon. The termination polymorphism can be detected with *Mnl*I and has an allele frequency of about 5%. The presence of the stop mutation phenotype (GG) appears to offer a protective effect toward CAD through the lowering of triglyceride and raising the HDL cholesterol concentrations *(43).* The CAD incidence of the heterozygous (GC) and stop polymorphism (GG) was lower than for the wild-type (CC) phenotype (OR: 0.38; 95% CI: 0.19–0.81) among 189 Japanese subjects. In contrast, a European study of 125 subjects showed no difference in C477G polymorphism between these CAD and control groups *(44).* Clearly, the number of subjects are too low at this time to make any definitive conclusions about the C477G polymorphism for lipoprotein lipase.

A SNP substitution at nucleotide 9 of exon 2 produces an amino acid change from aspartic acid to asparagine. The allele frequency is rather low at about 1–5% *(44)*. For this reason, this polymorphism has not yet been widely examined in population studies. Another A→G SNP has been identified in nucleotide 291 of exon 6, resulting in the substitution of an asparagine for a serine. The allele frequency of the Asn291Ser substitution is about 2.5%. Preliminary studies have suggested that the heterozygous carriers had increased triglycerides and decreased HDL cholesterol and a predisposition toward ischemic heart disease in women (OR; 1.89; 95% CI: 1.19–3.01) *(45)*.

4.2. PARAOXONASE Paraoxonase (PON) is a 44-kDa calcium-dependent ester hydrolase glycoprotein that is associated with apoA-I and located on the surface of HDL. The enzyme is encoded in the PON1 gene, which is part of a multigene family (PON2 and PON3) with a high degree of sequence homology. Currently, no protein product has been identified for PON2 and PON3 genes. It functions to inhibit the oxidation of LDL by preventing the formation of lipid peroxides and is, therefore, thought to be an important protective factor for atherosclerosis and CAD *(46)*. The gene that is most frequently studied is PON1, where there are two common polymorphisms. A change of a G→A results in an amino acid substitution of a glutamine (Q allele) to an arginine (R) at codon 192. A second polymorphism exists at codon 55, where T→A results in a substitution of a leucine (L allele) for a methionine (M). On the PON2 gene, there is a cysteine→serine substitution at codon 311. The allele frequency for the R and M polymorphism is 30% and 50%, respectively *(47)*. Among Asians, the 192R allele is higher (approx 50%) and the 55L allele is lower (10%) *(48)*. In contrast, among blacks, the 192R is lower (10%) than whites and the 55L allele is higher (73%) *(49)*. These genes exhibit linkage equilibrium *(50)*. The allele frequency for the 331S allele is about 75% *(49)*. Polymorphisms can also be found in the promoter region of PON1, such as a C→T substitution at –108 *(51)* and a C→G substitution at –907 *(52)*.

The polymorphism of the PON1 gene produces enzymes that have different reactivities toward natural and synthetic substrates. The 192 wild-type allele is more reactive for exogenous nerve gas substrates such as sarin and soman, whereas the R allele is more reactive toward paraoxon and oxon analogs *(46)*. The opposite occurs for the 55 polymorphism, where the wild-type L allele has the highest reactivity toward paraoxone.

There have been several clinical studies that have attempted to correlate the 192 and 55 PON1 polymorphism to cardiac diseases. The data published thus far are contradictory. No association was observed for either polymorphism among Europeans *(53,54);* however, there have been reports of an association among Asians for the 192R allele *(48,55,56)*. A meta-analysis has suggested that the QR and RR alleles are more prevalent in CHD than controls (OR: 1.44, 95% CI: 1.17–1.77) *(57)*. There have fewer studies conducted for the 311S polymorphism on the PON2 gene. Two studies have suggested a positive correlation with this allele and CAD *(58,59)*.

Because the association between paraoxonase polymorphisms have not produced consistent results, some investigators have suggested that the paraoxonase enzyme activity might be more important than the genotype. In one study, the paraoxonase activity of 417 angiographically proven CAD at 123 nmol/min was substantially lower than the 282 controls at 215 nmol/min *(57)*. In this study, there was no association between PON1 polymorphism and CAD incidence. These authors suggest that there are other factors in addition to the genotype that determine the activity of the final enzyme product.

5. NADH/NADPH OXIDASE

Superoxide (O_2•) induces oxidative stress and is an important factor in the development of atherosclerosis and coronary artery disease. O_2• is produced in vascular tissues by NADH/NADPH oxidase. High activity of this enzyme might be equated with increased free-radical production and a higher incidence of CAD. A major component of NADH/NADPH oxidase is the $p22^{phox}$ protein, which is part of a membrane-bound heterodimer and thought to be involved with heme binding. Other subunits of this enzyme include $p47^{phox}$, $p67^{phox}$, $gp91^{phox}$, and Rac-2 proteins. One study showed that the $p22^{phox}$ subunit was upregulated in atherosclerotic coronary arteries vs nonatherosclerotic arteries to suggest a role of this enzyme in the pathophysiology of CAD *(60)*.

A polymorphism exists in the $p22^{phox}$ protein, where a C→T substitution at codon 242 results in the substitution of tyrosine for a histidine at residue 72. There are significant racial differences in the frequency of the 242T allele. Among Caucasians, the frequency is 35% *(61),* much lower than among Asians at 13% *(62)*. The presence of the 242T allele was associated with significantly reduced vascular NAD(P)H activity of the enzyme (30%) relative to the wild-type *(63)*. One would then presume that the 242T allele would offer some protective effect against the development of CAD.

Studies attempting to correlate the presence of the 242T allele have been discordant. In a study of 402 CAD Japanese patients and controls, the incidence of the 242T (TT and CT) allele was significantly lower than the homozygote 242C wild type (OR: 0.49; 95% CI: 0.28–0.87), suggesting a protective effect for this polymorphism *(62)*. However, different results were observed among Caucasians, where the 242T allele was either not associated with CAD *(64)* or was associated with progession of atherosclerosis, as measured angiographically by a loss in mean minimum coronary artery lumen diameter *(65)*. One explanation for this apparent discordance is an upregulation of antioxidative defenses in Caucasians, despite the production of a defective enzyme.

5.1. ENDOTHELIAL NITRIC OXIDE SYNTHASE Nitric oxide (NO) is an important vasodilating factor that is released together with antioxidants from the endothelium under conditions of laminar arterial blood flow and high shear stress *(66)*. NO also inhibits platelet activation and leukocyte adhesion. Thus, it functions to protect the vessel from the development of atherosclerosis. Under conditions of disturbed blood flow and low shear stress, vasodilators and antioxidant are inhibited, and inflammation and apoptosis are stimulated, making the artery more prone to atherosclerotic lesions. Depressed NO concentrations have been found in early atherosclerosis and overt coronary artery disease. NO is synthesized from l-arginine via a family of nitric oxide synthases (NOS), of which there are at least three isoenzymes: inducible, constitutive neuronal, and

constitutive endothelial NOS(eNOS). eNOS is encoded by the NOS3 gene on chromosome 7. Polymorphisms in NOS3 might result in reduced or impaired function of the corresponding synthase.

A well-studied polymorphism occurs at position 894, where a G→T substitution results in a replacement of a glutamine to aspartic acid at position 298, and is thought to be proximal to the active site of the enzyme *(67)*. The frequency for the 298T allele among Caucasians is 30% *(67)* and about half this amount among Asian *(68)*. The 298T allele is associated with a decreased NO concentration, as measured by its metabolites (NO_2^- and NO_3^-) *(69)*. Studies on the association of the 298T allele to CAD have shown an association with CAD *(67,68,70)*; for example, Colombo et al. produced an odds ratio of 2.8 (95% CI: 1.2–6.8) for the 298TT genotype on 315 patients and controls *(70)*. However, like so many other genes, others have contradicted this conclusion *(71,72)*.

A VNTR polymorphism exists in intro 4, where the usual five copies of a 27-bp repeat (allele b) is replaced with four copies (a). The 4a frequency among Caucasians is 15% *(73)* and 10% among Asians *(74)*. Most studies published to date, however, have failed to find an association with this polymorphism and the presence of CAD among Caucasians *(73,75)* and Asians *(74,76)*. A less frequent polymorphism in the NOS3 gene is the –786 T→C substitution in the 5′ flanking promoter region. There appears to be a linkage between this polymorphism and the 4a VTNR polymorphism. Early studies have shown some association with coronary artery spasm *(74)* and early coronary artery disease *(77)*.

5.2. APO E GENOTYPES The apoproteins are the protein components of lipoproteins. There are five major apoproteins, labeled A–C, apo E, and apo (a), with subclasses that exist for many of these. Apoproteins function to activate enzymes in lipoprotein metabolism, maintain the structural integrity of the lipoprotein complex, and facilitate the uptake of lipoproteins into cells through surface receptors. Apolipoprotein E is a 34-kDa glycoprotein found in all major lipoproteins except LDL. Apo E plays an important role in the transportation and metabolism of triglycerides, chylomicrons, and VLDL remnants. There are three major isoforms of Apo E, designated E2, E3, and E4, and are encoded by the apo E gene, located on chromosome 19, to produce the ε2, ε3, and ε4 alleles. The ε2 contains a cysteine in amino residues 112 and 158, the ε3 contains a cysteine and arginine in these positions, and the ε4 contains an arginine in both locations. The presence of the ε4 genotype is associated with the presence of Alzheimer's disease *(78)*. The frequency of the ε2 allele is greatly dependent on the population. In one study of southern Europeans, it was 8%, ε3 was about 83%, and ε4 was about 9% *(79)*. The frequency of the ε4 allele is higher among African blacks.

Studies have shown that individuals with the ε2 alleles have the lowest LDL concentration and highest apo E, whereas those with the ε4 alleles have the reverse. It is possible that the apo E4 protein binds to the LDL receptor thereby interfering with the clearance of LDL from the circulation. Because LDL is recognized as a risk factor for CAD, studies have been conducted to determine if apo E genotypes are linked. A meta-analysis conducted by Wilson et al. combining the results of 9 published observational studies in men ($n = 1971$ cases) and 4 studies ($n = 181$ cases) in women *(80)*. The presence of the ε4 allele and presence of coronary heart disease produced an odds ratio of 1.38 (95% CI: 1.22–1.57) and 1.26 (95% CI: 1.13–1.41), respectively. In contrast, Frikke-Schmidt et al. found that the ε4 genotype was significantly associated with CHD for men ($n = 693$ cases) but not women ($n = 247$ cases) *(81)*.

Polymorphisms in other parts of the apolipoprotein E gene have been identified (i.e., within the promoter region). These include a A→T change at position –491, C→T change at position –427, and a G→T change at –219 *(82)*. In a study of 1245 AMI case and controls, the –219T polymorphism was associated with increased risk for AMI (OR: 1.29; 95% CI: 1.09–1.52). This polymorphism was not linked with the other polymorphisms in the promoter gene or with the ε alleles.

5.3. METHYLENE TETRAHYDROFOLATE REDUCTASE Homocysteine is formed from the metabolism of methionine, as shown in Fig. 2. Homocysteine is metabolized back to methionine with vitamins B6 and B12 and folate as cofactors, or transsulfurated to cysteine. High concentrations of homocsyteine have been implicated as a risk factor for a number of diseases such as peripheral vascular disease, stroke, and CAD. A meta-analysis was conducted examining the relationship of plasma homocysteine and development of CAD *(83)*. Of 14 case-control studies involving over 4700 patients, the odds ratio for development of CAD was 1.7 (95% CI: 1.5–1.9). Clinical trials are being conducted to determine if lowering homocysteine concentrations will result in a reduced risk for CAD. Commercial assays for homocysteine are available and are routinely used in clinical practice. The upper limit of the normal range for homocysteine in plasma is 15 μmol/L. However, a lower concentration that is within the normal range (e.g., 12 μmol/L) might be a more appropriate limit for low or no disease risk.

A major cause of hyperhomocysteinemia is dietary deficiency of vitamins B6 and B12 and folate. As such, the United States has recently introduced folate fortification in grain in an attempt to reduce the prevalence of dietary-induced hyperhomocysteinemia. Folate supplementation is inexpensive and has no side effects. The genetic basis of hyperhomocysteinemia has been known for many years and was the principal basis behind the hypothesis that abnormal homocysteine metabolism led to premature atherothrombotic events. Rare causes of homocystinuria include genetic deficiencies of methionine synthase and cystathionine β-synthase, important enzymes in the degradation of homocysteine. Pathologic changes within the coronary and cerebral arteries in children with congenital homocystinuria (i.e., intimal damage, hyperplasia of smooth muscles, lipoprotein aggregates, etc.) are identical to that found in adult atherosclerotic plaques.

A severe deficiency in methylenetetrahydrofolate reductase (MTHFR) can also cause hyperhomocysteinemia and homocystinuria. MTHFR is a homodimer of 77-kDa subunits and is a key enzyme in the folate cycle that generates 5-methyltetrahydrofolate, the substrate for methionine synthase (Fig. 2). The severe MTHFR deficiency is rare with only a few dozen cases described worldwide. At least 18 different point mutations have been demonstrated to produce these severe forms of MTFHR deficiency.

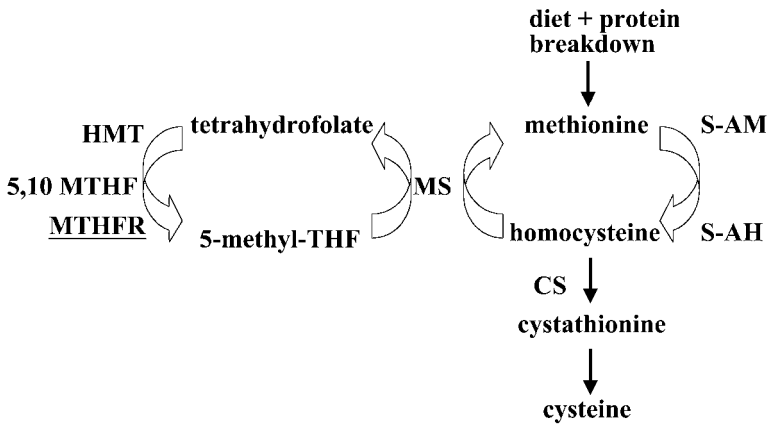


Fig. 2. Metabolic pathway for homocysteine. Methionine from the diet and metabolism of proteins is converted to *S*-adensylmethionine (S-AM), *S*-adenosylhomocysteine (S-AH), and then to homocysteine. With 5-methyl-tetrahydrofolate (5-THF) as a substrate, homocysteine is then remethylated back to methionine and tetrahydrofolate by methionine synthase (MS). Using vitamins B6 and B2 as cofactors, respectively, THF remethylated to 5-THF by serine–glycine hydroxymethyltransferase (HMT) and methylenetrahydrofolate reductase (MTHFR), respectively. Polymorphisms in the MTHFR gene is a subject of many clinical investigations. In an alternate pathway, homocysteine is also converted to cystathionine by cystathionine β-synthase (CS) and then to cysteine.

There are five common polymorphisms in MTHFR. The most common and widely studied of these is a C→ T substitution in nucleotide 677 of exon 4. This polymorphism has an allele frequency of about 35% and homozygosity frequencies of 10–15%. The homozygous genotype is associated with <20% of the normal MTHFR activity and thermolability in lymphocyte extracts resulting in low plasma folate concentrations and a mild hyperhomocysteinemia. In the study by Dunn et al., subjects under 50 yr old had a mean homocysteine concentration that varied from 15.8 to 15.9 for the homozygote wild type and heterozygote, which increased to 21.4 μmol/L for the homozygous C677T variant ($p < 0.0001$) *(84).*

There have been many studies that have attempted to correlate MTHFR polymorphism to CAD, with conflicting results. A meta-analysis of 10 studies on 5644 patients with CAD produced an odds ratio of 1.30 (95% CI: 1.11–1.52) for CAD and the presence of the 677T variant *(3).* This association was confirmed on a larger meta-analysis of 40 studies: 11,162 CAD cases and 12,758 controls (OR: 1.15, 95% CI: 1.05–1.28) *(85).* Among patients with renal failure, the 677T variant was associated with a significantly higher incidence of cardiovascular disease and homocysteine concentrations than patients without such a history *(86).* No association has been observed for this MTHFR mutation and ischemic stroke *(3,87).* A major limitation in many of these studies is the failure to consider plasma folate concentrations as a confounding variable. It might be possible that a reduced MTHFR activity can be compensated for by the presence of high plasma folate potentially normalizing the homocysteine concentration. The correlation of MTFHR TT polymorphmism with CAD might be even more significant if the studied population had been folate deficient.

6. MARKERS FOR HEART FAILURE

6.1. ANGIOTENSIN-CONVERTING ENZYME The angiotensin-converting enzyme (ACE) is a critical enzyme in the rennin–angiotensin–aldosterone (RAA) axis. Under the action of renin, angiotensinogen is first degraded to the decapeptide angiotensinI and then to the octapeptide angiotensinII by ACE. Angiotensin II is further degraded into a heptapeptide angiotensin III by angiotensinase. When there are periods of blood and fluid loss, the RAA axis is essential in maintaining blood pressure by stimulating vasoconstriction (via angiotensinII), sodium plasma volume retention and potassium loss, and inactivating bradykinin (vasodilator). Prolonged stimulation of the RAA axis leads to cardiovascular remodeling and has been implicated in the pathogenesis of neointimal hyperplasia. The use of ACE inhibitors has been shown to reduce atherosclerosis in animal models *(88).*

There is a common polymorphism in the angiotensin-converting enzyme gene in intron 16 involving the insertion (I) or deletion (D) of a 287-bp *alu* repeat sequence. The allele frequency for the insertion is 45%. The concentration of plasma ACE is increased in patients with the DD vs the II genotypes (e.g., 494 vs 299 ng/mL, respectively) *(89).* The first study that correlated the presence of the DD genotype to ACS was published in 1992 on 1343 cases and controls *(90).* Subsequently, a meta-analysis was conducted combining the results of 15 studies on 8873 cases and control *(91).* The cumulative meta-analysis produced an odds ratio of 1.26 (95% CI: 1.15–1.39) for the DD allele.

The accuracy of these early studies was questioned by Shanmugan et al., who showed that the D allele is preferentially amplified in heterozygote subjects *(92).* These investigators suggested the use of primers that specifically recognizes the insertion-specific sequence. More recent articles that have used the preferred primers have not demonstrated a relationship between ACE polymorphism and presence of CAD *(93).*

Given the role of the RAA axis on stimulating the growth of smooth muscles, ACE polymorphisms have been examined as a contributor to restenosis after angioplasty *(94).* These authors suggest that the pathophysiology of post-PTCA restenosis is related to smooth muscle cell migration and remodeling rather than proliferation. In contrast, stent placement might be more related to the growth effects of ACE. The study of Amant et al.

showed an inverse relationship between the number of D alleles present and the minimum lumen diameter (MLD) at 6 mo after successful stent implantation *(95)*. These observations would be significant if the MDL can be modified by use of ACE inhibitors such as quinapril. Unfortunately, the trials conducted to date have produced widely conflicting results. Although one trial showed beneficial effects of this drug in increasing MLD *(96),* another showed no effect *(97)*. Still another found that the D allele produced a detrimental effect *(98)*, whereas a fourth study suggested that the ACE I homozygotes and not the D genotype exhibited beneficial effects on MLD *(99)*. Thus, the role of ACE inhibition and ACE polymorphism after stent placement remains unclear at this time.

Because the RAA system is important in the pathophysiology of CHF, the effect of ACE polymorphisms has also been studied to determine if there is a correlation to CHF. Like results for CAD, the studies conducted to date are conflicting. Candy et al. correlated ACE genotypes with left ventricular systolic performance as measured by ejection fraction using both echocardiography and radionuclide ventriculography *(100)*. McNamara reported higher survival rates for CHF patients who were not treated with β-blockers *(101)*. However, Montgomery was unable to find an association of ACE polymorphism in 99 patients with idiopathic dilated cariomyopathy *(102)*. One explanation for the lack of an association between the DD polymorphism was suggested by Spruth et al. *(103),* who found no difference in mRNA expression among the DD, ID, and II genotypes. Finding genetic markers for patients with CHF might be more difficult than for patients with ACS because there might be more heterogeniety in the pathophysiology of CHF.

The ACE D/I polymorphism has also been examined in patients with atrial fibrillation. In a pilot study, the II genotype was a significant risk factor for atrial fibrillation (OR: 3.2; 95% CI: 1.3–7.8) among 138 patients with dilated cardiomyopathy *(104)*. This interesting observation needs to be confirmed with larger clinical trials.

Given the importance of ACE inhibitors for the treatment of patients with essential hypertension, there have been a number of clinical studies conducted to determine if the D/I polymorphism in the ACE gene correlates to susceptibility to antihypertensive therapy. Such a finding could pave the way for the determination of genotypes on an individual basis prior to the selection of a particular drug regimen. A review of studies was recently conducted by Niu et al. *(105)*. In some studies, there was no significant differences in genotypes and clinical measures such as left ventricular modeling *(106)*. It is possible that these studies were underpowered to find a significant difference *(107)*. In other studies, favorable outcomes (e.g., reduction in left ventricular hypertrophy) were found for individuals with the DD genotype *(108),* whereas other studies suggested more favorable outcomes (e.g., reduction in diastolic blood pressure) for the II genotype *(109)*. As essential hypertension is a multifactorial disease, more careful selection of hypertensive subjects will be necessary to sort out the discordances that currently exist in the literature. In patients suffering AMI, the success of post-AMI treatment with ACE inhibitors was evaluated according to ACE D/I polymorphism, using a reduction in left ventricular size or function as the outcome measures. Unfortunately, there was no association between genotype and outcomes among 265 patients to suggest a pharcogenomic role in improving clinical efficacy *(110)*.

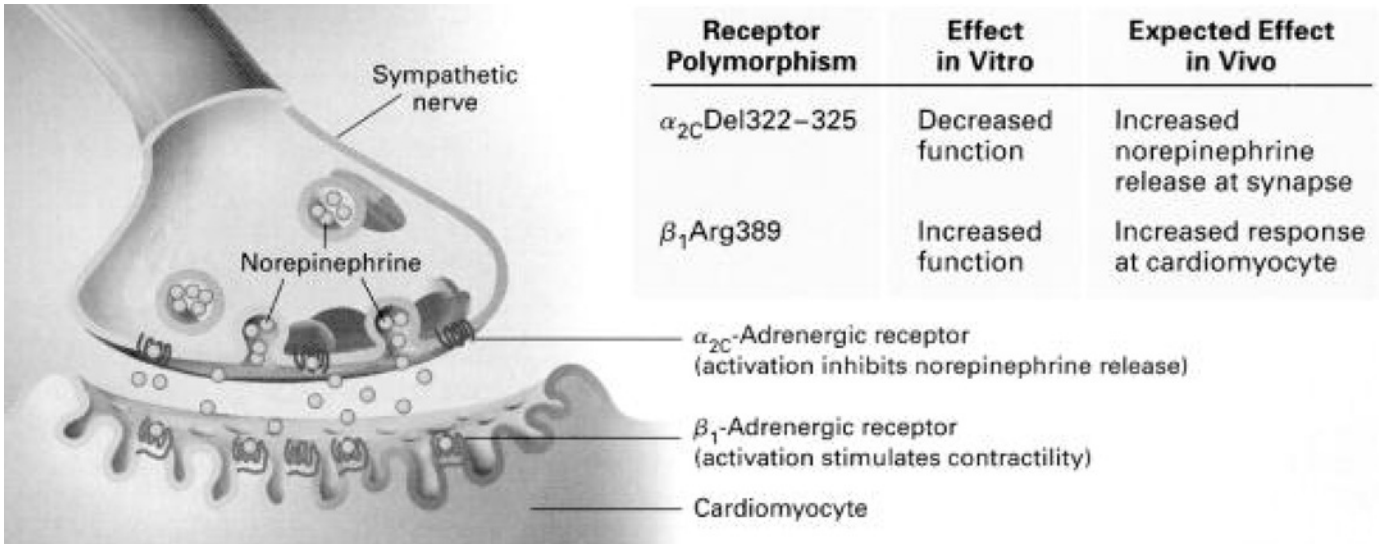
A polymorphism also exists in the gene for angiotensin receptor type 1, where there is a change from an A→T substitution in nucleotide 1166. This mutation might be linked to increased prevalence of hypertension and coronary vasoconstriction. A early study suggested that this polymorphism was linked to the ACE D/I polymorphism; that is, the combination of the DD and 1166CC phenotype produced a significant association for AMI (OR: 3.95; 95% CI: 1.26–12.4) *(111)*. This association has not been confirmed in more recent studies *(112)*. In a study in which angiotensin II was given to volunteers, investigators concluded that the A1166C polymorphism does not have an effect on the actions of angiotensin II *(113)*.

6.2 α- AND β-ADRENERGIC RECEPTORS Cardiac inotropy and chronotropy are modulated in part by α- and β-adrenergic receptors expressed in the human heart. Both α and β receptors are subdivided into subtypes. The α_1 subtypes are localized in the postsynaptic junctions regulating catecholamine uptake, whereas the α_2 subtypes are found in presynaptic junctions and regulating catecholamine release *(114)*. There are four α_1– ($\alpha_{1A–D}$), and three α_2– ($\alpha_{2A–C}$) subtypes. Although several polymorphisms have been described in these subtypes, only a few have been examined in conjunction with CAD. A polymorphism in the α_{1A} subtype has been described where a cysteine is substituted for arginine in position 492. Although there are differences in the frequency of this mutation among Caucasians and African-Americans, there was no association of this polymorphism with essential hypertension *(115)*. There are two nucleotide substitution polymorphisms in the α_{1B} adrenergic receptor at nucleotide positions 534 and 549 (but do not result in an amino acid substitution) and a C→G substitution in the promoter region of α_{2A}, but none of these polymorphisms were associated with resting and challenged blood pressures *(116)*.

A three-amino-acid deletion polymorphism in the α_{2B} receptor within a 12 glutamic acid repeat segment (positions 297–309) has been linked to a reduced basal metabolic rate and obesity *(117)* and flow-mediated dilatation of the brachial artery *(118)*. Both obesity and defects in endothelial function are known risk factors for CAD. There has been one study to date that attempted to link this polymorphism to cardiovascular disease. In 912 Finnish men, the homozygous deletion genotype was associated with an odds ratio of 2.2 (95% CI: 1.1–4.4) for development of an ACS compared to the heterozygote and wild-type genotypes *(119)*. There was no association in this study between these genotypes and hypertension.

A four-amino-acid deletion polymorphism exists in the α_{2C} receptor located at positions 322–325, resulting in the loss of agonist-mediated receptor function in transfected cells *(120)*. The loss of inhibitory function results in the release of epinephrine and overstimulation of the cardiovasculature *(121)*.

More research in the role of polymorphism in hypertension and heart disease has been conducted with the β-adrenergic receptors. These receptors are important in activating adenylate cyclase to produce cyclic AMP (second messenger) that augment myocardial contractility. Patients with heart failure have an



Receptor Polymorphism	Effect in Vitro	Expected Effect in Vivo
α_{2C}Del322–325	Decreased function	Increased norepinephrine release at synapse
β_1Arg389	Increased function	Increased response at cardiomyocyte

Fig. 3. Putative mechanism for the role of α- and β- adrenergic receptor polymorphisms as risk factors for heart failure. The α_{2c}-adrenergic receptor inhibits norepinephrine release. A deletion of amino acids 322–325 results in a loss of receptor function and overstimulation of catecholamines at synpaptic junctions. Epinephrine activates β_1-adrenergic receptors via adnenylate cylase within myocytes resulting in increased heart rate and force. The Arg389 polymorphism in the β_1-adrenergic gene results in increased coupling to adenylyl cylase and enhanced function. (Used with permission from ref. *125*.)

overstimulation of epinephrine, resulting in a downregulation of the β receptors. The use of β-blocker therapy improves clinical symptoms and cardiac output by slowing the heart rate and improving the rhythm. Under a situation of increased RAA axis stimulation, polymorphisms in the β receptors might predispose an individual to the development of heart failure.

There are many polymorphisms in the β_1-receptor gene that encode for a 478-amino-acid protein. Many of the polymorphisms, however, are silent and, thus, their significance toward development of CHF can be questioned *(122).* Two of these polymorphisms have been studied with regard to idiopathic dilated cardiomyopathy. One involves nucleotide 145, where there is a change from an A→G resulting in the substitution of serine for a glycine. The other is at nucleotide 1165, where there is a change from a C→G resulting in an arginine to a glycine substitution. The allele frequencies for the 145G and 1165C are 15% and 25%, respectively. These polymorphisms were studied in a cohort of patients with CHF. In two studies totaling 1063 cases and 1660 controls, the 1165G polymorphism was not associated with idiopathic dilated cardiomyopathy *(123–125).* In another study, Borjesson et al. found the 49G allele frequency was higher than for the 49S allele *(126).* These investigators also reported significant differences in survival rates with the mutation (OR: 2.34; 95% CI: 1.30–4.20). The N-terminal sequence where the 49S mutation resides might be important to fold the receptor within the membrane while not being the target for receptor activation. These observations may help explain why some individuals respond to β-blockade while others do not. Prospective pharmacogenomic trials for different β_1- and β_2-blockers will be the next logical step in this field.

Three common polymorphism in the β_2 receptor in amino acid position 16 substituting an arginine for a glycine, position 27 substituting a glutamine for a glutamic acid, and 164 substituting an arginine for an isoleucine. The allele frequencies for the respective polymorphisms are 61%, 43%, and 2%. The polymorphism in positions 16 and 27 have different susceptibility to agonist-induced downregulation. The 164 polymorphism exhibits a decreased affinity for β_2-adrenogeric angonists, and an uncoupling of receptors from the G_s protein. In volunteer subjects, the increase in heart rate and systolic blood pressure among those with the Thr164Ile polymorphism was less than for the wild type *(127).* In a study of 471 case and controls, Liggett et al. found no difference in the allele frequencies for CHF patients and controls *(128).* However, when the 1-yr outcomes were compared, patients with the 164 ile genotype had a higher death risk rate and need for cardiac transplantation than CHF patients with the wild-type mutation. In 140 cases and controls, Wieczorek et al. was able to find an increased incidence of the 164 ile genotype with CHF *(129).*

Recently, Small et al. correlated the combination of the adrenergic receptor polymorphism of α_{2C}-322–325 deletion and β_1-arg389gly substitution to risk of CHF in black subjects *(125).* The basis of their hypothesis is summarized in Fig. 3. On only 159 CHF patients and 189 matched controls, they reported odds ratios for the α_{2C} deletion polymorphism alone, which increased to an odds ratio of 5.65 (95% CI: 2.67–11.95) and 10.11 (95% CI: 2.11–48.53) when combined with the β_1-adrenergic receptor polymorphism. There was no correlation with the β_1 receptor alone with CHF and insufficient data to make any conclusions in Caucasians. These data exhibit the highest odd ratios for any polymorphism to cardiac disease published to date.

7. SUMMARY

Although many of these genes demonstrate a significant correlation with polymorphisms and the incidence of CAD or heart failure, the degree of risk is not particularly high in any case (odds ratio typically between 1 and 2). Most investigators

agree that the pathogenesis of these diseases is too complex and diverse to be explained on the basis of a SNP. Discrepancies in results of individual studies might be the result of inadequate numbers of subjects enrolled or heterogeneity in the sex and/or racial make of the population studied. However, several studies have shown that the presence of some mutations are linked to others and their effect on disease progression might be additive. Unfortunately, all of the studies published to date involve only a handful of gene mutations on a limited number of subjects enrolled. The costs for performing the DNA analysis prohibits the analysis for more than just of few of these genes at a time. This chapter discussed only a minority of the many polymorphisms that are known to exist among the various proteins and factors known that participate in the disease.

Genes that have been initially targeted are those proteins known to participate in the pathophysiology of cardiac diseases. It might be appropriate for future studies to concentration on the effect of polymorphisms in the success of therapy, both pharmacologic and surgical (e.g., PCI). This justifies the current interest in the genes that modify platelet function, ACE inhibitors, and β-blockers. The work should be further focused in therapies where there is variability in response and/or success (e.g., hypertension). A specific polymorphorism might be an important factor in the lack of success in some instances.

REFERENCES

1. Grimm, R. and Rettig, R. Association studies between the agniotensin-converting enzyme insertion/deletion polymorphism and hypertension: still interesting? *J. Hypertens,* **20:**1049–1051, 2002.
2. Jones, A. and Montgomery, H. The gly389arg beta-1 adrenoceptor polymorphism and cardiovascular disease: time for a rethink in the funding of genetic studies? *Eur. Heart J.* **23:**1071–1074, 2002.
3. Ridker, P. M., Hennekens, C. H., and Miletich, J. P. G20210A mutation in prothrombin gene and risk of myocardial infarction, stroke, and venous thrombosis in a large cohort of US men. *Circulation* **99:**999–1004, 2000.
4. Wu, A. H. B. and Tsongalis, G. J. Correlation of selected single nucleotide polymorphisms of coagulation factors to risk for cardiovascular disease. *Am. J. Cardiol*. **87:**1361–1366, 2001.
5. Meade, T. W., Mellows, S., Brozovic, M., et al. Haemostatic function and ischaemic heart disease: principal results of the Northwick Park Heart Study. *Lancet*. **2:**533–537, 1986.
6. Girelli, D., Russo, C., Ferraresi, P., et al. Polymorphisms in the Factor VII gene and the risk of myocardial infarction in patients with coronary artery disease. *N. Engl. J. Med*. **343:**774–780, 2000.
7. Fellowes, A. P., Brennan, S. O., and George, P. M. Identification and characterization of five new fibrinogen gene polymorphisms. *Ann. NY Acad. Sci.* **936:**536–541, 2001.
8. Green, F. R. Fibrinogen polymorphisms and atherothrombotic disease. *Ann. NY Acad. Sci.* **936:**549–559, 2001.
9. Doggen, C. J. M., Bertina, R. M., Cats, V. M., and Rosendaal, F. R. Fibrinogen polymmorphisms are not associated with the risk of myocardial infarction. *Brit. J. Haematol.* **110:**935–938, 2000.
10. de Maat, M. P., Kastelein, J. J., Jukema, J. W., et al. -455G/A polymorphism of the beta-fibrinogen gene is associated with the progression of coronary atherosclerosis in symptomatic men: proposed role for an acute-phase reaction pattern of fibrinogen. REGRESS group. *Arteriosclosis Thromb. Vasc. Biol*. **18:**265–271, 1998.
11. Ridker, P. M., Vaughan, D. E., Stampfer, M. J., Manson, J. E., and Hennekens, C. H. Endogenous tissue-type plasminogen activator and risk of myocardial infarction. *Lancet* **341:**1165–1168, 1993.
12. Juhan-Vague, I., Pyke, S. D., Alessi, M. C., Jespersen, J., Haverkate, F., and Thompson, S. G. Fibrinolytic factors and the risk of myocardial infarction or sudden death in patients with angina pectoris. ECAT Study Group. European Concerted Action on Thrombosis and Disabilities. *Circulation* **94:**2057–2063, 1996.
13. Ye, S., Green, F. R., Scarabin, P. Y., et al. The 4G/5G genetic polymorphism in the promoter of the plasminogen activator-1 (PAI-1) gene is associated with differences in plasma PAI-1 activity but not with risk of myocardial infarction in the ECTIM study. *Thromb. Haemost*. **74:**837–841, 1995.
14. Iacoviello, L., Burzotta, F., Di Castelnuovo, A., Zilo, F., Marchioli, R., and Donati, M. D. The 4G/5G polymorphism of PAI-1 promoter gene and the risk of myocardial infarction: a meta analysis. *Thromb. Haemost.* **80:**1029–1030, 1998.
15. Ossei-Gerning, N., Mansfield, M. W., Stickland M. H., Wilson, I. J., and Grant, P. Plasminogen activator inhibitor-1 promoter 4G/5G genotype and plasma levels in relation to a history of myocardial infarction in patients characterized by coronary angiography. *Arteriosclosis Thromb. Vasc. Biol.* **17:**33–37, 1997.
16. Mizuno, K., Satomura, K., Miyamoto, A., et al. Angioscopic evaluation of coronary-artery thrombi in acute coronary syndromes. *N. Engl. J. Med.* **326:**287–291, 1992.
17. George, J. N. Platelets. *Lancet* **355:**1531–1541, 2000.
18. Weiss, E. J., Bray, P. F., Tayback, M., et al. A polymorphism of a platelet glycoprotein receptor as an inherited risk factor for coronary thrombosis. *N. Engl. J. Med.* **334:**1090–1094, 1996.
19. Ridker, P. M., Hennekens, C. H., Schmitz, C., Stampfer, M. J., and Lindpaintner, K. $PI^{A1/A2}$ polymorphism of platelet glycoprotein IIIa and risks of myocardial infarction, stroke, and venous thrombosis. *Lancet* **349:**385–388, 1997.
20. Rinder, C., Mathew, J. P., Rinder, H. M., et al. for the Multicenter Study of Perioperative Ischemia Research Group. Platelet PI^{A2} polymorphism and platelet activation are associated with increased troponin I release after cardiopulmonary bypass. *Anesthesiology* **97:**1118–1122, 2002.
21. Lemmer, J. H. Clinical experience in coronary bypass surgery for abciximab-treated patients. *Ann. Thorac. Surg.* **70:**S33–S37, 2000.
22. Juergens, C. P., Yeung, A. C., and Oesterle, S. N. Routine platelet transfusion in patients undergoing emergency coronary bypass surgery after receiving abciximab. *Am. J. Cardiol.* **80:**74–75, 1997.
23. Wheeler, G. L., Braden, G. A., Bray, P. F., Marciniak, S. J., Mascelli, M. A., and Sane, D. C. Reduced inhibition by abciximab in platelets with the PI^{A1} polymorphism. *Am. Heart. J.* **143:**76–82, 2002.
24. Laule, M., Cascorbi, I., Stangl, V., et al. A1/A2 polymorphism of glycoprotein IIIa and association with excess procedural risk for coronary catheter interventions: a case-controlled study. *Lancet*. **353:**708–712, 1999.
25. Kastrati, A., Koch, W., Gawaz, M., et al. P1A polymorphism of glycoprotein IIIa and risk of adverse events after coronary stent placement. *J. Am. Coll. Cardiol.* **36:**84–89, 2000.
26. Bottiger, C., Kastrati, A., Koch, W., et al. Polymorphism of platelet glycoprotein IIb and risk of thrombosis and restenosis after coronary stent placement. *Am. J. Cardiol*. **84:**987–991, 1999.
27. Furihata, K., Nugent, D. J., and Kunicki, T. J. Influence of platelet collagen receptor polymorphisms on risk for arterial thrombosis. *Arch. Patho. of Lab. Med.* **126:**305–309, 2002.
28. Kunicki, T. J. The influence of platelet collagen receptor polymorphisms in hemostasis and thrombotic disease. *Arteriosclosis Thromb. Vasc. Biol.* **22:**14–20, 2002.
29. Santoso, S., Kunicki, T. J., Kroll, H., Haberbosch, W., and Gardemann, A. Association of the platelet glycoprotein Ia C_{807} T gene polymorphism with nonfatal myocardial infarction in younger patients. *Blood* **93:**2449–2453, 1999.
30. Mikkelsson, J., Perola, M., Penttila, A., and Karhunen, P. J. Platelet collagen receptor GPIa (C807T/HPA-5) haplotype is not associated with an increased risk of fatal coronary events in middle-aged men. *Atherosclosis* **165:**111–118, 2002.
31. Von Beckerath, N., Koch, W., Mehilli, J., Bottiger, C., Schomig, A., and Kasatrati, A. Glycoprotein Ia gene C807T polymorphism and risk for major adverse cardiac events within the first 30 days after coronary artery stenting. *Blood* **95:**3297–3301, 2000.
32. Kandzari, D. E. and Goldschmidt-Clermont, P. J. Platelet polymorphisms and ischemic heart disease: moving beyond traditional risk factors. *J. Am. Coll. Cardiol.* **38:**1028–1032, 2001.

33. Corral, J., Gonzalez-Conejero, R., Lozano, M. L., Rivera, J., and Vicente, V. New alleles of the platelet glycoprotein Ibα gene. *Br. J. Haematol.* **103:**997–1003, 1998.
34. Mikkelsson. J., Perola, M., Penttila, A., and Karhunen, P. J. Platelet glycoprotein Ibα HPA-2 Met/VNTR B haplotype as a genetic predictor of myocardial infarction and sudden cardiac death. *Circulation* **104:**876–880, 2001.
35. Meisel, C., Afshar-Kharghan, V., Cascorbi, I., et al. sequence polymorphism of platelet glycoprotein Ibα as a risk factor for coronary artery disease and catheter interventions. *J. Am. Coll. Cardiol.* **38:**1023–1027, 2001.
36. Corral, J., Lozano, M. L., Gonzalez-Conejero, R., et al. common polymorphism flanking the ATG initiator codon of GPIb alpha does not affect expression and is not a major risk factor for arterial thrombosis. *Thromb. Haemost.* **83:**23–28, 2000.
37. Croft, S. A., Samani, N. J., Teare, M. D., et al. Novel platelet membrane glycoprotein VI dimnorphism is a risk factor for myocardial infarction. *Circulation* **104:**1459–1463, 2001.
38. Gerdes, C., Gerdes, L. U., Hansen, P. S., and Faergeman, O. Polymorphisms in the lipoprotein lipase gene and their associations with plasma lipid concentrations in 40-year old danish men. *Circulation* **92:**1765–1769, 1995.
39. Larson, I., Hoffmann, M. M., Ordovas, J. M., Schaefer, E. J., Marz, W., and Kreuzer, J. The lipoprotein lipase *Hin*dIII polymorphism: association with total cholesterol and LDL-cholesterol, but not with HDL and triglycerides in 342 females. *Clin. Chem.* **45:**963–968, 1999.
40. Wang, X. T., McCredie, R. M., and Wilcken, D. E. L. Common DNA polymorphisms at the lipoprotein lipase gene: association with severity of coronary artery disease and diabetes. *Circulation* **93:**1339–1345, 1996.
41. Gambino, R., Scaglione, L., Alemanno, N., Pagano, G., and Cassader, M. Human lipoprotein lipase *Hin*dIII polymorphism in young patients with myocardial infarction. *Metabolism* **48:**1157–1161, 1999.
42. Anderson, J. L., King, G. J., Bair, T. L., et al. Association of lipoprotein gene polymorphisms with coronary artery disease. *J. Am. Coll. Cardiol.* **33:**1013–1020, 1999.
43. Sawano, M., Watanabe, Y., Ohmura, H., et al. Potentially protective effects of the Ser 447-ter mutation of the lipoprotein lipase gene against the development of coronary artery disease in Japanese subjects via a beneficial lipid profile. *Jpn. Circ. J.* **65:**310–314, 2001.
44. Raslova, K., Smolkova, B., Vohnout, B., Gasparovic, J., and Frohlich, J. J. Risk factors for atherosclerosis in survivors of myocardial infarction and their spouses: comparison to controls without personal and family history of atherosclerosis. *Metabolism* **50:**24–29, 2001.
45. Wittrup, H. H., Tybjaerg-Hansen, A., Abildgaard, S., Steffensen, R., Schnohr, P., and Nordestgaard, B. G. A common substitution (Asn291Ser) in lipoprotein lipase is associated with increased risk of ischemic heart disease. *J. Clin. Invest.* **99:**1606–1613, 1997.
46. Mackness, M. I., Mackness, B., Durrington, P. N., et al. Paraoxonase and coronary heart disease. *Curr. Opin. Lipid* **9:**319–324, 1998.
47. Garin, M. C. B., James, R.W., Dussoix, P., et al. Paraoxonase polymorphism met-leu54 is associated with modified serum concentrations of the enzyme: a possible link between the paraoxonase gene and increased risk of cardiovascular disease in diabetes. *J. Clin. Invest.* **99:**62–66, 1997.
48. Zama, T., Murata, M., Matsubara, Y., et al. A192 arg variant of the human paraoxonase (HUMPONA) gene polymorphism is associated with an increase risk for coronary aratery disease in the Japanese. *Arteriosclosis Thromb. Vasc. Biol.* **17:**3565–3569, 1997.
49. Chen, Q., Reis, S. E., Kammerer, C. M., et al. for the WISE Study Group. Association between the severity of angiographic coronary artery disease and paraoxonase gene polymorphisms in the National Heart Lung, and Blood Institute-sponsored Women's Ischemic Syndrome Evaluation (WISE) Study. *Am. J. Hum. Genet.* **72:**13–22, 2003.
50. Sanghera, D. K., Saha, N., Aston, C. E., and Kamboh, M. I. Genetic polymorphism of paraoxonase and the risk of coronary heart disease. *Arteriosclosis Thromb. Vasc. Biol.* **17:**1067–1073, 1997.
51. James, R. W., Leviev, I., Ruiz, J., et al. Promoter polymorphism T(–107)C of the paroxonase PON1 gene is a risk factor for coronary heart disease in type 2 diabetic patients. *Diabetes* **49:**1390–1393, 2000.
52. Levie, I., Poirier, O., Nicaud, V., et al. High expressor paraoxonase PON1 gene promoter polymorphisms are associated with reduced risk of vascular disease in younger coronary patients. *Atherosclosis* **161:**463–467, 2002.
53. Ferre, N., Tous, M., Paul, A., et al. Paraoxonase gln-arg(192) and leu-met(55) gene polymorphisms and enzyme activity in a population with a low rate of coronary heart disease. *Clin. Biochem.* **35:**197–203, 2002.
54. Heijmans, B. T., Westendorp, R. G., Lagaay, A. M., Knook, D. L., Kluft, C., and Slagboom, P. E. Common paraoxonase gene variants, mortality risk and fatal cardiovascular events in elderly subjects. *Atherosclo*sis **149:**91–97, 2000.
55. Osei-Hyiaman, D., Hou, L., Menghai, F., Zhihin, R., Zhiming, Z., and Kano, K. Coronary artery disease risk in Chinese type 2 diabetics: is there a role for paraoxonase 1 gene (Q192R) polymorphism? *Eur. J. Endocrinol.* **144:**639–644, 2001.
56. Odawara, M., Tachi, Y., and Yamashita, K. Paraoxonase polymorphism (gln192-arg) is associated with coronary heart disease in Japanese noninsulin-dependent diabetes mellitus. *J. Clin. Endocrinol. Metab.* **82:**2257–2260, 1997.
57. Mackness, B., Davies, G. K., Turkie, W., et al. Paraoxonase status in coronary heart disease: are activity and concentration more important than genotype? *Arteriosclosis Thromb. Vasc. Biol.* **21:**1451–1457, 2001.
58. Sanghere, D. K., Saha, N., Aston, C. E., and Kamboh, M. I. Genetic polymorphism of paraoxonase and the risk of coronary heart disease. *Arteriosclosis Thromb. Vasc. Biol.* **17:**1067–1073, 1997.
59. Pan, J. P., Lai, S.T., Chiang, S. C., Chou, S. C., and Chiang, A. N. The risk of coronary artery disease in population of Taiwan is associated with cys-ser 311 polymorphism of human paraoxonase (PON)-2 gene. *Chin. Med. J.* **65:**415–421, 2002.
60. Azumi, H., Inoue, N., Takeshita, S., et al. Expression of NADH/NADPH oxidase $p22^{phox}$ in human coronary arteries. *Circulation* **100:**1494–1498, 1999.
61. Cai, H., Duarte, N., Wilcken, D. E., and Wang, X. L. NADH/NADPH oxidase p22 phox C242T polymorphism and coronary artery disease in the Australian population. *Eur. J. Clin. Invest.* **29:**744–748, 1999.
62. Inoue, N., Kawashima, S., Kanazawa, K., Yamada, S., Akita, H., and Yokoyama, M. Polymorphism of the NADH/NADPH oxidase p22 phox gene in patients with coronary artery disease. *Circulation* **97:**135–137, 1998.
63. Guzik, T., West, N. E., Black, E., et al. Functional effect of the C242T polymorphism in the NAD(P)H oxidase p22phox gene on vascular superoxide production in atherosclerosis. *Circulation* **102:**1744–1747, 2000.
64. Li, A., Prasad, A., Mincemoyer, R., et al. Relationship of the C242T p22phox gene polymorphism to angiographic coronary artery disease and endothelial function. *Am. J. Med. Genet.* **86:**57–61, 1999.
65. Cahilly, C., Ballantyne, C. M., Lim, D. S., Gotto, A. M., and Ali, J. A variant of $p22^{phox}$, involved in generation of reactive oxygen species in the vessel wall, is associated with progression of coronary atherosclerosis. *Circ. Res.* **86:**391–395, 2000.
66. Kinlay, S., Libby, P., and Ganz, P. Endothelial function and coronary artery disease. *Curr. Opin. Lipid.* **12:**383–389, 2001.
67. Hingorani, A. D., Liang, C. F., Fatibene, J., et al. A common variant of the endothelial nitric oxide synthase (glu298asp) is a major risk factor for coronary artery disease in the UK. *Circulation* **100:**1515–1520, 1999.
68. Hibi, K., Ishigami, T., Tamura, K., et al. Endothelial nitric oxide synthase gene polymorphism and acute myocardial infarction. *Hypertension* **32:**521–526, 1998.
69. Sofowora, G., Dishy, V., Xie, H. G., et al. In-vivo effects of glu298asp endothelial nitric oxide synthase polymorphism. *Pharmacogenesis* **11:**809–814, 2001.
70. Colombo, M. G., Andreassi, M. G., Paradossi, U., et al. Evidence for association of a common variant of the endothelial nitric oxide synthase gene (glu^{298}→asp) polymorphism to the presence, extent, and severity of coronary artery disease. *Heart* **87:**525–528, 2002.

71. Nassar, B. A., Bevin, L. D., Johnstone, D. E., et al. Relationship of the glu298asp polymorphism of the endothelial nitric oxide synthase gene and early-onset coronary artery disease. *Am. Heart J.* **142:**586–589, 2001.
72. Wang, C. L., Hsu, L. A., Ko, Y. S., Ko, Y. L., and Lee, Y. H. Lack of association between the glu298asp variant of the endothelia nitric oxide synthase gene and the risk of coronary artery disease among Taiwanese. *J. Form. Med. Assoc.* **100:**736–740, 2001.
73. Granath, B., Taylor, R. R., van Bockxmeer, F. M., and Mamotte, C. D. S. Lack of evidence for association between endothelial nitric oxide synthase gene polymorphisms and coronary artery disease in the Australian Caucasian population. *J. Cardiovasc. Risk* **8:**235–241, 2001.
74. Yoshimura, M., Yasue, H., Nakayama, M., et al. Genetic risk factors for coronary artery spasm: significance of endothelial nitric oxide synthase gene T-789→C and missense Glu298asp variants. *J. Invest. Med.* **48:**367–374, 2000.
75. Sigusch, H. H., Surber, R., Lehmann, M. H., et al. Lack of association between 27-bp repeat polymorphoism in intron 4 of the endothelial nitric oxide synthase gene and the risk of coronary artery disease. *Scand. J. Clin. Lab. Invest.* **60:** 229–235, 2000.
76. Nakagami, H., Ikeda, U., Maeda, Y., et al. Coronary artery disease and endothelial nitric oxide synthase and angiotensin-converting enzyme gene polymorphisms. *J. Thromb. Thrombol.* **8:**191–195, 1999.
77. Alvarez, R., Gonzalez, P., Batalla, A., et al. Association between the NOS3 (-786 T/C) and the ACE (I/D) DNA genotypes and early coronary artery disease. *Nitric Oxide* **5:**343–348, 2001.
78. Higgins, G. A., Large, C. H., Rupniak, H. T., and Barnes, J. C. Apolipoprotein E and Alzheimer's disease: a review of recent studies. *Pharmacol. Biochem. Behav.* **56:**675–685, 1997.
79. Corbo, R. M., Vilardo, T., Ruggeri, M., Gemma, A.T., and Scacchi, R. Apolipoprotein E genotype and plasma levels in coronary artery disease. A case-control study in the Italian population. *Clin. Biochem.* **32:**217–222, 1999.
80. Wilson, P.W., Schaefer, E. J., Larson, M.G., and Ordovas, J. M. Apolipoprotein E alleles and risk of coronary disease. A meta-analysis. *Arteriosclosis Thromb. Vasc. Biol.* **16:**1250–1255, 1996.
81. Frikke-Schmidt, R., Tybjaerg-Hansen, A., Steffensen, R., Jensen, G., and Nordestgaard, B. G. Apolipoprotein E genotype: epsilon32 women are protected while epsilon43 and epsilon44 men are susceptible to ischemic heart disease. *J. Am. Coll. Cardiol.* **35:** 1192–1199, 2000.
82. Lambert, J. C., Brousseau, T., Defosse, V.,et al. Independent association of an APOE gene promoter polymorphism with increased risk of myocardial infarction and decrease APOE plasma concentrations—the ECTIM study. *Hum. Mol. Genet.* **9:**57–61, 2000.
83. Boushey, C.J., Beresford, S. A., Omen, G. S., and Motulsky, A. G. A quantitative assessment of plasma homocysteine as a risk factor for vascular disease. *JAMA* **274:**1049–1057, 1995.
84. Dunn, J., Title, L. M., Bata I., et al. Relation of a common mutation in methylenetetrahydrofolate reductase to plasma homocysteine and early onset coronary artery disease. *Clin. Biochem.* **31:**95–100, 1998.
85. Klerk, M., Verhoef, P., Clarke, R., Bloom, H. J., Kok, F. J., and Schouten, E. G. MTHFR 677C→polymorphism and risk of coronary heart disease: a meta-analysis. *JAMA* **288:**2023–2031, 2002.
86. Morimoto, K., Haneda, T., Okamoto, K., Ishida, H., and Kikuchi, K. Methylenetetrahydrofolate reductase gene polymorphism, hyperhomocysteinemia, and cardiovascular diseases in chronic hemodialysis patients. *Nephron* **90:**43–50, 2002.
87. Kelly, P. J., Rosand, J., Kistler, J. P., et al. Homocysteine, MTHFR 677C→T polymorphism, and risk of ischemic stroke: results of a meta-analysis. *Neurology* **59:**529–536, 2002.
88. Shuh, J. R., Bleh, D. J., Frierdich, G. E., McMahon, E. G., and Blaine, E. H. Differential effects of rennin–angiogensin system blockade on atherogenesis in cholesterol-fed rabbits. *J. Clin. Invest.* **981:**1453–1458, 1993.
89. Rigat, B., Hubert, C., Alhenc-Gelas, F., Cambien, F., Corvol, P., and Soubrier, F. An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels. *J. Clin. Invest.* **86:**1343–1346, 1990.
90. Cambien, F., Poirier, O., Lecerf, L., et al. Deletion polymorphism in the gene for angiotensin-converting enzyme is a potential risk factor for myocardial infarction. *Nature* **359:**641–643, 1992.
91. Samani, N. J., Thompson, J. R., O'Toole, L., Channer, K., and Woods, K. L. A meta-analysis of the association of the deletion allele of the angiotensin-converting enzyme gene with myocardial infarction. *Circulation* **94:**708–712, 1996.
92. Shanmugan, V., Sell, K.W., and Saha, B.K. Mistyping ACE heterozygotes. *PCR Methods* applic. **3:**120–121, 1993.
93. Rice, G. I., Foy, C. A., Grant, P. J. Angiotensin converting enzyme and angiotensin II type 1-receptor gene polymorisms and risk of ischaemic heart disease. *Cardiovasc. Res.* **41:**746–753, 1999.
94. Zee, R.Y. L, Fernandez-Ortiz, A., Macaya, C., Pintor, E., Lindpaintner, K., and Fernandez-Cruz, A. ACE D/I polymorphism and incidence of post-PTCA restenosis: a prospective, angiography-based evaluation. *Hypertension* **37:**851–855, 2001.
95. Amant,C., Bauters, C., Bodart, J. C., et al. D allele of the angiotensin I-converting enzyme is a major risk factor for restenosis after coronary stenting. *Circulation* **96:**56–60, 1997.
96. Okumura, K., Sone, T., Kondo, J.,et al. Quinapril prevents restenosis after coronary stenting in patients with angiotensin-converting enzyme D allele. *Circ. J.* **66:**311–316, 2002.
97. Koch, W., Kastrati, A., Mehilli, J., Bottiger, C., von Beckerath, N., and Schomig, A. Insertion/deletion polymorphis of the angiotensin I-converting enzyme gene is not associated with restenosis after coronary stent placement. *Circulation* **102:**197–202, 2000.
98. Meurice, T., Bauters, C., Hermant, X., et al. Effect of ACE inhibitors on angiographic restenosis after coronary stenting (PARIS): a randomised, double-blind, placebo-controlled trial. *Lancet* **357:**1321–1324, 2001.
99. Toyofyuku, M., Imazu, M., Sumii, K., et al. Influence of angiotensinogen M253T gene polymorphism and an angiotensin converting enzyme inhibitor on restenosis after percutaneous coronary intervention. *Atherosclosis* **160:**339–344, 2002.
100. Candy, G. P., Skudicky, D., Mueller, U. K., et al. Association of left ventricular systolic performance and cavity size with angiotensin-converting enzyme genotype in idiopathic dilated cardiopathy. *Am. J. Cardiol.* **83:**740–744, 1999.
101. McNamara, D. M., Holubkov, R., Janosko, K., et al. Pharmacogenetic interactions between β-blocker therapy and the angiotensin-converting enzyme deletion polymorphism in patients with congestive heart failure. *Circulation* **103:**1644–1648, 2001.
102. Montgomery, H. E., Keeling, P. J., Goldman, J. H., Humphries, S. E., Talmud, P. J., and McKenna, W. J. Lack of association between the insertion/deletion polymorphism of the angioten-converting enzyme gene and idiopathic dilated cardiomyopathy. *J. Am. Coll. Cardiol.* **25:**1627–1631, 1995.
103. Spruth, E., Zurbrugg, H. R., Warnecke, C., et al. Expresion of ACE mRNA in the human atrial myocardium is not dependent on left ventricular function, ACE inhibitor therapy, or the ACE I/D genotype. *J. Mol. Med.* **77:**804–810, 1999.
104. Ogimoto, A., Hamada, M., Nakura, J., Miki, T., and Hiwada, K. Relation between angiotensin-converting enzyme II genotype and atrial fibrillation in Japanese patients with hypertrophic cardiomyopathy. *J. Hum. Genet.* **47:**184–189, 2002.
105. Niu, T., Chen, X., and Xu, X. Angiotensin converting enzyme gene insertion/deletion polymorphism and cardiovascular disease. Therapeutic inplications. *Drugs* **62:**977–993, 2002.
106. Matsubara, M., Suzuki, M., Fujiwara, T., et al. Angiotensin-converting enzyme I/D polymorphism and hypertension: the Ohasama study. *J. Hypertens.* **20:**1121–1216, 2002.
107. Grimm, R., and Rettig, R. Association studies between the angiotensin-converting enzyme insertion/deletion polymorphism and hypertension: still interesting? *J. Hypertens.* **20:**1049–1051, 2002.
108. Sasaki, M., Oki, T., and Iuchi, A. Relationship between the angiotensin converting enzyme gene polymorphism and the effects of enalapril on left ventricular hypertrophy and impaired diastolic filling in essential hypertension: M-mode and pulsed Doppler echocardiographic studies. *J Hypertens.* **14:**1403–1408, 1996.
109. Ohmichi, N., Iwai, N., Uchida, Y., Shichiri, G., Nakamura, Y., and Kinoshita, M. Relationship between the response to the angiotensin

converting enzyme inhibitor imidapril and the angiotensin converting enzyme genotype. *Am. J. Hypertens.* **10:**951–955, 1997.
110. Zee, R.Y., Solomon, S. D., Ajani, U.A., Pfeffer, M.A., Lindpaintner, K., for the HEART Investigators. A prospective evaluation of the angiotensin-converting enzyme D/I polymorphism and left ventricular remodeling in the "Healing and Early After Reducing Therapy" study. *Clin.Genet.* **61:**21–25, 2002.
111. Tiret, L., Bonnardeaux, A., Poirier, O., et al. Synergistic effects of angiotensin-converting enzyme and angiotensin-II type 1 receptor gene polymorphisms on risk of myocardial infarction. *Lancet* **344:**910–913, 1994.
112. Batalla, A., Alvarez, R., Reguero, J. R., et al. Synergistic effect between apolipoprotein E and angiotensinogen gene polymorphisms in the risk for early myocardial infarction. *Clin. Chem.* **46:**1910–1915, 1999.
113. Hilgers, K. F., Langenfeld, M. R.W., Schlaich, M., Veelken, R., and Schmieder, R. E. 1166 A/C polymorphism of the angiotensin II type 1 receptor gene and the response to short-term infusion of angiotensin II. *Circulation* **100:**1394–1399, 1999.
114. Veglio, F., Tayebati, S. K., Schiavone, D., et al. α_1-Adrenergic receptor subtypes in peripheral blood lymphocytes of essential hypertensives. *J. Hypertens.* **19:**1847–1854, 2001.
115. Xie, H. G., Kim, R. B., Stein, C.M., Gainer, J.V, Brown, N.J., and Wood, A.J. Alpha 1A-adrenergic receptor polymorphism: association with ethnicity but not essential hypertension. Pharmacogenetics **9:**651–656, 1999.
116. McCaffrey, J.M., Pogue-Geile, M.F., Ferrell, R. E., Petro, N., and Manuck, S. B. Variability within α and β-adrenoreceptor genes as a predictor of cardiovascular function at rest and in response to mental challenge. *J. Hypertens.* **20:**1105–1114, 2002.
117. Heinonen, P., Loulu, M., Pesonen, U., et al. Identification of a three-amino acid deletion in the α_{2B}-adrenergic receptor that is associated with reduced basal metabolic rate in obese subjects. *J. Clin. Endocrinol. Metab.* **84:**2429–2433, 1999.
118. Heinonen, P., Jartti, L., Jarvisalo, M. J., et al. Deletion polymorphism in the α_{2B}-adrenergic receptor gene is associated with flow-mediated dilatation of the brachial artery. *Clin. Sci.* **103:**517–524, 2002.
119. Snapir, A., Heinonen, P., Tuomainen, T. P., et al. An insertion/detetion polymorphismis a novel genetic risk factor for acute coronary events. *J. Am. Coll. Cardiol.* **37:**1516–1522, 2001.
120. Small, K. M., Forbes, S. l., Rahman, F. F., Bridges, K. M., and Ligett, S. B. A four amino acid deletion polymorphism in the third intracellular loop of the human alpha 2C-adrenergic receptor confers impaired coupling to multiple effectors. *J. Biol. Chem.* **275:** 23059–23064, 2000.
121. Hajjar, R.J., and MacRae, C. A. Adrenergic-receptor polymorphisms and heart failure. *N. Engl. J. Med*. **347:**1196–1199, 2002.
122. Podlowski, S., Wenzel, K., Luther, H.P., et al. β_1-Adrenoceptor gene variations: a role in idiopathic dilated cardiomyopathy? *J. Mol. Med.* **78:**87–93, 2000.
123. Tesson, F., Charron, P., Peuchmaurd. M., et al. Characterization of unique genetic variant in the β_1-adrenoceptor gene and evaluation of its role in idiopathic dilated cardiomyopathy. *J. Mol. Cell Cardiol.* **31:**1025–1032, 1999.
124. White, H.L., Maqbool, A., McMahon, A. D., et al. An evaluation of the beta-1 adrenergic receptor Arg389Gly polymorphism in individuals at risk of coronary events. A WOSCOPS substudy. *Eur. Heart. J.* **23:**1087–1092, 2002.
125. Small, K. M., Wagoner, L. E., Levin, A. B., Kardia, S. L.R., and Liggett, S.B. Synergistic polymorphisms of β_1- and α_{1c}-adrenergic receptors and the risk of congestive heart failure. *N. Engl. J. Med.* **347:**1135–1142, 2002.
126. Borjesson, M., Magnusson, Y., Hjalmarson, A., and Andersson, B. A novel polymorphism in the gene coding for the beta$_1$-adrenergic receptor associated with survival in patients with heart failure. *Eur. Heart J.* **21:**1853–1858, 2000.
127. Brodde, O. E., Rainer, B., Tellkamp, R., Radke, J., Stefan, D., and Insel, P.A. Blunted cardiac responses to receptor activation in subjects with Thr164Ile β_2-adrenoceptors. *Circulation* **103:**1048–1050, 2001.
128. Liggett, S. B., Wagoner, L. E., Craft, L. L., et al. The Ile164 β2-adrenergic receptor polymorphism adversely affects the outcome of congestive heart failure. *J. Clin. Invest.* **102:**1534–1539, 1998.
129. Wieczorek, S. J., Wu, A. H. B., and Tsongalis, G. Unpublished data.

26 Molecular Diagnostics in Coagulation

ENRIQUE BALLESTEROS

1. INTRODUCTION

Coagulation testing has evolved from the use of the manual tilt-tube clotting time in the early 1900's to the use of molecular diagnostics today. Over the years, the measurement of a clot endpoint has been the basis of testing. Automated instruments have replaced the manual visual methods and allowed for precise measurements of the clot endpoint. In addition, immunologic and chromogenic methodology has also been adapted into the coagulation laboratory, therefore providing an additional antigenic and enzymatic perspective. Today, most modern automated coagulation instruments offer a single test platform that incorporates optic, immunologic, and chromogenic methods. The arrival of molecular diagnostics now adds another dimension for the evaluation of hemostatic defects.

Historically, coagulation testing was confined mostly to the screening or work-up of patients with bleeding disorders. The availability of routine laboratory tests for patients with thrombosis has slowly increased over the years and the demand for this testing continues to grow. Molecular diagnostics has found an important role in this area of coagulation disorders. DNA-based tests are available for detection of the Factor V Leiden mutation, the Prothrombin 20120A mutation, and the methylenetetrahydrofolate reductase mutation.

At the same time, the significance of molecular testing for coagulation defects associated with thrombosis and the impact on patient care continues to be controversial. Physicians from different specialties are struggling with the questions surrounding the availability of a larger test "menu": whom, when, and what to test? *(1)*. Given the tendency to practice predictive medicine and the resulting increased screening of specific subgroups of patients for specific genetic information, it is anticipated that the role of molecular diagnostics for coagulation disorders will continue to increase with time *(2)*.

In this chapter, a brief review of the procoagulant and anticoagulant systems will be followed by a general description of the molecular defects underlying various coagulation disorders. This, in turn, will provide a background for a review on the molecular genetic testing available in coagulation, particularly for the hereditary hypercoagulable states. Finally, the clinical significance of testing for Factor V Leiden, Prothrombin G20210A, and the methylenetetrahydrofolate reductase mutations will be discussed.

2. OVERVIEW OF COAGULATION

2.1. GENERAL CONCEPTS The components of coagulation include the blood vessels, platelets, coagulation factors and cofactors, and the fibrinolytic proteins. As a result, hemostasis is a delicate balance between surface proteins on vessels/subendothelium, surface glycoproteins on platelets, procoagulant and anticoagulant proteins, and fibrinolytic proteins *(3)*. The interaction among these components results in both fibrin clot formation and dissolution at the site of injury. Hereditary or acquired defects in any of these components can result in bleeding or thrombosis.

Hemostasis is generally viewed as two linked processes: primary and secondary hemostasis *(4)*. Primary hemostasis, through platelet adhesion, activation, and aggregation, represents the platelet response to damaged endothelium and, at the same time, provides a template for the coagulation cascade. The coagulation cascade of proenzymes and cofactors is, in turn, referred to as secondary hemostasis. Together, these two processes form the basis of clot formation.

2.2. PRIMARY HEMOSTASIS With damage to a blood vessel, the exposed collagen binds circulating Von Willebrand factor (VWF), which, in turn, functions as a "glue" between the subendothelium and the circulating platelets *(5)*. Platelets adhere to the damaged vessel through the platelet receptor for VWF, glycoprotein Ib (GpIb), followed by platelet activation with release of ADP and serotonin, among others, from their storage granules. Platelet activation also results in expression of the surface receptor glycoprotein IIb/IIIa (GpIIb/IIIa), which, together with fibrinogen and other proteins, allows for platelet aggregation. Meanwhile, the presence of phosphaditylserine on the surface of the platelet plug provides a negatively charged phospholipid surface, necessary for the subsequent formation of the fibrin clot *(4)*.

2.3. SECONDARY HEMOSTASIS Secondary hemostasis is characterized by the formation of a fibrin meshwork, which serves to reinforce the platelet plug. The system is comprised of a coagulation cascade of proenzymes and it's activated forms,

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

along with cofactors and calcium *(4)*. The procoagulants include the contact system proteins (factors XII and XI, high-molecular-weight kallikrein and kininogen), the vitamin K-dependent proteins (factors II, VII, IX, X), fibrinogen, the cofactors V and VIII, and factor XIII.

In vitro observations led to the traditional concept of two separate pathways involved in the generation of thrombin and the subsequent fibrin clot: the intrinsic and extrinsic pathways *(3)*. In the intrinsic pathway, the components are found "inside the blood," best demonstrated when whole blood in a glass tube is left to clot by itself. In contrast, in the extrinsic pathway, clotting is initiated by tissue factor, which might originate from one of several sources located "outside the blood." The extrinsic pathway is generally considered the physiologic pathway in vivo while the intrinsic pathway provides a reinforcing mechanism *(6)*.

The coagulation cascade is initiated when tissue factor binds to the small amount of circulating activated factor VII and this complex, in turn, activates both factors IX and X *(3)*. Of note, the in vivo activation of factor IX by factor VII has replaced the traditional view of distinct intrinsic and extrinsic pathways. In the presence of cofactor VIII, activated factor IX will further activate factor X. Likewise, activated factor X, in the presence of cofactor Va, will subsequently convert prothrombin to thrombin, followed by the conversion of fibrinogen to fibrin.

Thrombin plays a central role in *procoagulation:* activating platelets, converting fibrinogen to fibrin, and providing positive feedback for further activation of cofactors V and VIII and factor XI. At the same time, thrombin also acts as an *anticoagulant* by activating the protein C system. Thrombin loses its procoagulant activity by binding to thrombomodulin present in most endothelial cell surfaces, followed by activation of the natural anticoagulant, protein C *(6)*.

2.4. THE NATURAL ANTICOAGULANT SYSTEMS Physiologic anticoagulant system(s) help keep thrombin formation in check, thus avoiding abnormal clot propagation. Tissue factor pathway inhibitor (TFPI), serine protease inhibitors (so-called serpins), the protein C system, and the fibrinolytic system are all necessary for the regulation of thrombin and fibrin formation *(4)*.

The protein C system, through protein C, protein S, and thrombomodulin, plays a major anticoagulant role by inactivation of cofactors V and VIII *(7)*. Activated protein C, in the presence of phospholipid and the cofactor protein S, degrades specific peptide bonds at positions Arg 506 and Arg 306 in the B-domain of the factor V heavy chain.

The fibrinolytic system includes plasminogen, plasminogen activators and activator inhibitors, plasmin, and α_2-antiplasmin *(8)*. Plasminogen is activated mainly by tissue plasminogen activator (tPA) to form plasmin, which then degrades the clot. Plasminogen activation is confined to the fibrin clot, limiting the overall effect of fibrinolysis to the clot itself.

In sum, the physiologic balance of procoagulation and anticoagulation results in controlled thrombin formation limited to the site of vascular injury. A number of mutations and polymorphisms in the genes coding for these various proteins might result in increase risk of bleeding or thrombosis. A few of these genetic defects are now evaluable through molecular testing.

3. MOLECULAR DEFECTS UNDERLYING COAGULATION DISORDERS

3.1. POLYMORPHISMS AND MUTATIONS Homologous areas of the genomic DNA contain variations within their nucleotide sequence. When this sequence variation has a greater than 1% frequency in the population, it is referred to as a polymorphism *(9)*. These polymorphisms can occur in genes and gene-related sequences or noncoding extragenic DNA. Alleles represent polymorphisms within a single gene *(10)*. When the sequence variation has a population frequency of less than 1%, this might be referred to as a mutation; however, some mutations might be present with >1% frequency (i.e., Factor V Leiden). Many different polymorphisms/mutations form the molecular basis of risk factors associated with coagulation disorders.

Different types of genetic polymorphisms, including single-nucleotide polymorphisms (SNPs), deletions, duplications, and insertions, have been variably associated with different coagulation disorders *(11–15)*. With several genes of coagulation proteins now completely sequenced, a number of SNPs have been described, including point mutations in exons, introns, or regulatory regions *(11)*. These mutations can result in loss of function (e.g., protein C deficiency) or gain of function (e.g., Prothrombin mutation 20210A) (*11,14*).

Not suprisingly, given their level of genetic complexity, coagulation disorders might be phenotypically heterogeneous. Many different mutations and polymorphisms might affect the same gene; there might be silent mutations or frameshift mutations, the latter typically associated with a severe phenotypic abnormality *(11)*. Levels of coagulation factors might vary depending on the type of polymorphism or mutation *(14,16)*. This degree of genetic complexity remains a challenge for the molecular diagnostics laboratory.

The genetic complexity in coagulation disorders is further magnified by the interaction between different genes (*17*) and between genes and the environment *(18)*. Although mutations in several different genes might be associated with hereditary thrombophilia, the age of onset, severity of the disease, and penetrance are highly variable. Factor V Leiden has a high prevalence in selected ethnic groups; approx 3–7% of the white population of northern European ancestry are heterozygous carriers, but the disease has a relatively low penetrance *(19)*. In contrast, a condition with a low prevalence such as the homozygous state for protein C deficiency might be associated with a severe clinical syndrome, purpura fulminans, in the newborn *(20)*. In the vast majority of adult patients, venous thrombosis most likely represents the interaction of environmental and genetic factors *(18)*. The interaction with acquired risk factors compounds the risk of venous thromboembolism; a single gene defect alone might have a limited role in thrombosis, but in combination with known acquired risk factors, the two circumstances together might significantly increase the risk for disease *(18)*.

3.2. MOLECULAR DEFECTS IN BLEEDING DISORDERS The clinical use of molecular diagnostics apply mainly to thrombophilic disorders, however, the molecular defects underlying both bleeding and thrombotic disorders are similar.

Furthermore, there may be a role for molecular genetic testing in bleeding disorders in the near future. A comprehensive review of the molecular abnormalities in bleeding disorders is beyond the scope of this chapter; however, a few comments on the most common hereditary bleeding disorders will follow.

More than 200 mutations with diverse effects have been described for the factor VIII gene *(21)*. Of interest, specific mutations are associated with particular disease phenotypes: insertions in the factor VIII gene are generally associated with severe hemophilia A *(22)*, whereas a missense mutation has been associated with the development of a factor VIII inhibitor (*23*). Likewise, many different mutations have been described for factor IX *(24)*. In contrast to the large gene defects typically seen in hemophilia A, small gene defects are usually associated with factor IX deficiency (hemophilia B). Different polymorphisms in the factor IX gene result in variable clinical disorders: hemophilia B Leiden is associated with a bleeding disorder that improves with age *(25)*, whereas the ALA-10 factor IX polymorphism results in profuse bleeding after initiating oral anticoagulant therapy (*26*).

Likewise, many different mutations and polymorphisms have been described for the common hereditary bleeding disorder, Von Willebrand disease (VWD). Interestingly, most of these known mutations are associated with type II VWD, whereas the genetic defects for the more common type I VWD are largely unknown *(27)*. A molecular-based classification of vWD has been proposed by Schneppenheim et al. *(28)*.

Missense and nonsense mutations, deletions, both small and large, insertions, and frameshift mutations, are described for factors VII (*29*), X (*30*), XI *(31)*, XII *(32),*and XIII *(33)* and fibrinogen *(34)*. The effects of these mutations are variable. For example, large deletions and frameshift mutations will likely result in an abnormal factor X, whereas point mutations that lead to replacement of a single amino acid are often associated with measurable factor X and mild disease *(30)*. Similarly, a number of different mutations in the fibrinogen gene(s) can result in afibrinogenemia and dysfibrinogenemias, with a variable clinical phenotype ranging from asymptomatic individuals to relatively severe bleeding or thrombotic disorders *(34)*.

3.3. MOLECULAR DEFECTS UNDERLYING THROMBOTIC DISORDERS As previously described, several natural anticoagulants protect the host from excessive clot formation. Various genetic defects translate into well-described hereditary thrombophilias (Table 1).

Table 1
Most Common Hereditary Thrombophilias

Factor V Leiden
Prothrombin G20210A mutation
Antithrombin deficiency
Protein C deficiency
Protein S deficiency

Activated protein C (APC) serves as an anticoagulant by inactivating factor V at amino acid positions Arg 506, Arg 306, and Arg 679, located in the B-domain of the factor V protein. Dahlback et al. *(35)* in 1993 described an abnormal phenotype associated with thrombosis and characterized by activated protein C resistance using clot-based assays. Subsequently, a mutation associated with this phenotype was described in the factor V gene, a guanine to adenine substitution at nucleotide 1691 of exon 10 *(36)*. This point mutation results in the substitution of the amino acid arginine by glutamine at residue 506 and is the underlying genetic defect for most cases associated with the APC resistance phenotype *(36)*. Although the mutated factor V protein can still be inactivated by APC, it does so at a much slower rate, thereby permitting continued formation of thrombin and subsequent clot formation. The Factor V Leiden mutation does not affect the *procoagulant* activity of factor V.

Of note, other mutations in addition to the FVL R506Q might be associated with the APC resistance phenotype. Factor V Cambridge *(37)* and factor V Hong Kong *(38)* result in APC resistance; however, in contrast to Factor V Leiden, their association with thrombosis is controversial *(15,39)*. Similarly, a mutation in exon 13 (4070A→G) of the factor V gene together with other sequence variations, referred to as the HR2 haplotype, is also associated with the APC resistance phenotype and possibly with thrombosis *(40,41)*.

The Prothrombin G20210A mutation is the second most common hereditary defect associated with thrombophilia. In 1996, Poort et al. *(42)* described a mutation in the prothrombin gene associated with a gain of function and presumably increased prothrombin levels. The mutation is localized in the 3′ untranslated region of the gene and consists of a guanine to adenine transition at nucleotide 20210, resulting in increased efficiency of mRNA end formation (polyadenylation) *(42,43)*. Not all individuals with elevated prothrombin levels have the mutation, and the exact mechanism of how the mutation results in thrombophilia is undetermined. It is unclear whether the G20210A mutation itself is the cause for the elevated prothrombin levels or whether the defect is in linkage disequilibrium with the responsible mutation *(15)*.

The association between antithrombin deficiency and thrombophilia was first described in 1965 *(44)*. Missense, frameshift, and deletion-type mutations can occur either in the gene loci corresponding to the so-called reactive site or the heparin-binding site of the protein. Mutations in the heparin-binding site might lack an association with thrombosis *(45)*.

Protein C is another important physiologic anticoagulant whose hereditary deficiency from loss of function mutations is an established risk factor for thrombosis in the heterozygous state, whereas the severe syndrome known as purpura fulminans is typically associated with homozygous inheritance with complete absence of protein C production *(20)*. More than 160 mutations, including missense and frameshift mutations, have been described for the protein C gene *(46)*. Interestingly, the protein C gene has a high number of CpG dinucleotides, which appear to be a hot spot for many of these nonrandom mutations *(11)*.

Similarly, greater than 100 mutations, including missense, frameshift insertions, and deletions, have been described for the protein S gene *(47)*. Adding to the complex protein S genetics, a pseudogene has also been described for this cofactor *(48)*.

Family studies have suggested an association between polymorphisms in the thrombomodulin gene and thrombosis *(12,49)*. However, mutations in both the gene and its promoter region have a variable association with arterial and venous

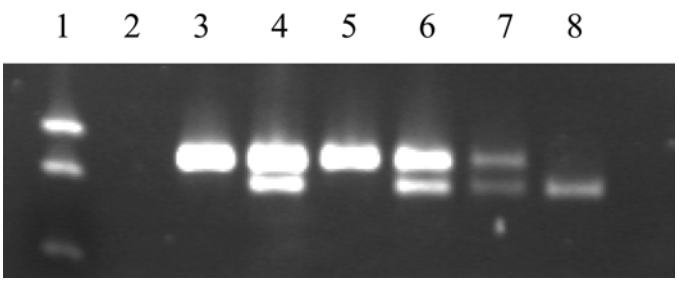


Fig. 1. Polyacrylamide gel electrophoresis of polymerase chain reaction–restriction fragment length polymorphism (PCR-RFLP) products for the C677T MTHFR polymorphism. Lane 1, molecular size markers; lane 2, blank control; lane 3, normal control; lane 4, heterozygous control; lane 5, normal patient; lanes 6 and 7, heterozygous patients; lane 8, homozygous patient.

thrombosis and, therefore, more studies are needed to determine its role as a risk factor *(15,50).*

Increased levels of procoagulant proteins, mainly factors VII and VIII and fibrinogen, have also been proposed as risk factors for thrombosis, particularly arterial disease *(12).* Coagulation protein levels might be influenced by variations in gene-promoter sequences and/or environmental factors *(15).* Although a relationship between high procoagulant levels and thrombosis is well documented, an association between polymorphisms in various coagulation factors and thrombosis itself has not been established *(12).* Specifically, the increased risk of thrombosis associated with elevated factor VIII levels has not been linked to a specific molecular defect of the factor VIII gene *(15,51).*

Although most of the defects associated with thrombosis involve either natural procoagulant or anticoagulant proteins, hereditary and acquired defects of the separate homocysteine metabolic pathway might also be associated with venous and arterial thrombosis *(52).* Hyperhomocysteinemia might be the result of acquired conditions such as Vit. B6/B12/folate deficiency and renal insufficiency, or genetic defects, including mutation of the methylenetetrahydrofolate reductase (MTHFR) gene. A C-T missense mutation in the MTHFR gene at nucleotide 677 creates a new restriction site for the restriction enzyme *Hin*f1 (Fig.1) *(53).* The mutation results in the amino acid substitution of alanine for valine and is associated with elevated levels of homocysteine (hyperhomocysteinemia). Similar to the other above-described genetic defects, more than one polymorphisms can affect the MTHFR gene *(54).* The precise molecular mechanisms for the elevated homocysteine levels are not known.

Finally, platelets are an integral part of the hemostatic process as described earlier. Polymorphisms in genes coding for the platelet surface proteins GpIb/IX and GpIIb/IIIa have been associated with arterial thrombotic disease *(12).* However, similar to other polymorphisms described for the coagulation proteins, more studies are needed to support the need for routine testing for these platelet polymorphisms (*See* also Chapter 25).

In summary, the understanding of the molecular basis of various coagulation defects continues to evolve and future studies will determine whether detection of these mutations/polymorphisms have a practical role in the laboratory work-up of bleeding and thrombotic disorders.

4. MOLECULAR TESTS IN THE CLINICAL COAGULATION LABORATORY

4.1. GENERAL ASPECTS The applications of molecular diagnostics in coagulation testing are confined mainly to patients with thrombosis, an area for which laboratory testing has increased steadily over the past years. The availability of molecular diagnostics for hereditary thrombophilic disorders depends on the underlying molecular defect. Based on their unique molecular abnormalities, the FVL mutation, Prothrombin G20210A mutation, and the MTHFR mutation all lend themselves to relatively straightforward molecular analysis. In contrast, other hereditary thrombophilic defects such as antithrombin, protein C, and protein S deficiencies are not typically assessed by molecular methods in the clinical laboratory given their many different mutations.

Molecular testing for the FVL, the Prothrombin G20210A mutation, and the MTHFR mutation have become readily available in many clinical laboratories. Several molecular methods are available and the choice of a particular method is based on the experience of the laboratory and the availability of infrastructure. Similar to other areas of a clinical laboratory, reagent cost and labor costs account for most of the expenses of a molecular pathology laboratory.

By far the most common molecular approach for the hereditary thrombophilias are polymerase chain reaction (PCR)-based assays. PCR is ideal for detection of point mutations or single-nucleotide polymorphisms (SNPs), including those associated with coagulation disorders. A high degree of concordance between laboratories shows that Factor V Leiden can be reliably detected by a variety of PCR-based assays *(55).*

EDTA-anticoagulated peripheral blood is collected from the patient and DNA is extracted from the white blood cells. Subsequently, in a typical PCR assay, the template DNA is mixed with a pair of oligonucleotide primers of varying length, specific for a target gene sequence. After the numerous cycles of denaturation, annealing, and polymerase extension (synthesis) in a thermocycler, a PCR product is obtained that can then be analyzed by one of several different methods *(56).* With multiplex PCR, more than one PCR product can be obtained in a single test run. Several manual and automated multiplexed PCR assays have been developed for simultaneous detection of Factor V Leiden, prothrombin 20210A mutation, and the MTHFR mutations *(57–59).*

The most common method used to analyze PCR products is restriction fragment length polymorphism (RFLP) analysis. Other molecular methods that include PCR-based and non-PCR-based assays, some of which eliminate the use of restriction digestion, have also been incorporated into the molecular diagnostics laboratory with improved cost-efficiency *(60–63).* Furthermore, the use of hybridization probes have become commonplace for analysis and monitoring of the PCR product *(56).* In response to a need for rapid, simple, and reliable diagnostic assays, sequence specific primers (PCR-SSP) *(60),* allele-specific oligonucleotide (ASO) hybridization *(55,56),* rapid-cycle PCR using the LightCycler instrument *(61),* and the

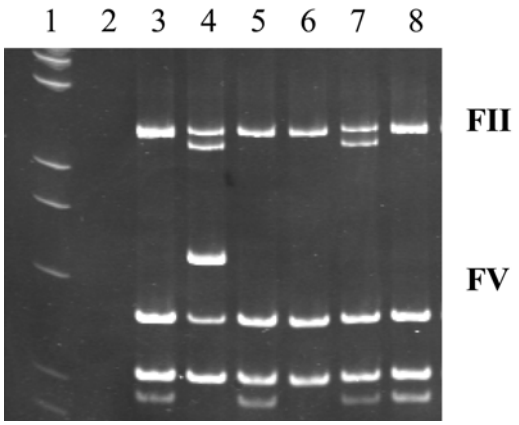


Fig. 2. Polyacrylamide gel electrophoresis of PCR-RFLP products for the Factor V Leiden (bottom) and prothrombin (FII) G20210A (top) polymorphisms. Lane 1, molecular size markers; lane 2, blank control; lane 3, normal control; lane 4, double heterozygous control; lanes 5, 6 and 8, normal patients; lane 7 heterozygous G2010A and normal FV Leiden patient.

Fig. 3. The Roche LightCycler real-time PCR instrument. For more information, see https://biochem.roche.com/lightcycler.

Nanochip technology based on electronic microarrays *(62)* have all been incorporated for molecular diagnosis in coagulation disorders.

4.2. RESTRICTION FRAGMENT LENGTH POLYMORPHISM The original molecular approach used for the detection of the FVL was PCR followed by RFLP. The FVL mutation abolishes a normal restriction enzyme site for *Mnl*I and creates a new restriction fragment pattern, allowing the mutant allele to be distinguished from the normal allele *(36)* (Fig. 2). The *Mnl*I fragments can be separated on agarose or polyarcylamide gels and visualized as wild-type or mutant bands. Of note, the *Mnl*I cleavage site might rarely be affected by other mutations that, although distinct from the FVL, could give a false-positive test with RFLP *(55)*.

If the inherited mutation does not result in change of the restriction enzyme site, a restriction site can also be introduced through site-directed mutagenesis. For the detection of the Prothrombin G20210A mutation, a mutagenic primer creates a new *Hin*dIII restriction site, allowing identification of the mutant allele *(42)* (Fig. 2). As with FVL testing, typically PCR is used to amplify the 3′ untranslated region of the Prothrombin gene surrounding the G20210A polymorphism, followed by restriction endonuclease digestion *(64)*.

4.3. PCR–SEQUENCE SPECIFIC PRIMERS A PCR reaction with sequence specific primers (PCR-SSP), without the need for postamplification restriction enzyme analysis, is also available for FVL analysis *(60)*. Using a sense primer complementary to both FV alleles along with either of two antisense allele-specific primers, each complementary to the normal and mutant allele, this simple and rapid PCR assay correlated 100% with the PCR-RFLP assay *(60)*.

4.4. ALLELE-SPECIFIC OLIGONUCLEOTIDE HYBRIDIZATION Allele-specific oligonucleotide (ASO) hybridization analysis is also available for FVL analysis. With ASO hybridization, both normal and mutant alleles are detected when the PCR-amplified products are spotted onto two membranes followed by hybridization with normal and mutant labeled probes *(55,56)*. This method is very useful when the SNP does not result in change of restriction enzyme recognition sites *(56)*.

4.5. LIGHTCYCLER RAPID PCR In the LightCycler (Roche) platform (Fig. 3), amplification of DNA in glass capillaries occurs under very rapid cycling conditions *(61)*. During PCR, probes labeled with fluorescein (FLU) and LC-Red640 emit fluorescence, that, depending on the degree of match between the probes and target DNA, will decrease when dissociation of the hybridized probes occurs under temperature control. A single-point mutation resulting in a nonperfect match will produce an earlier dissociation with a given temperature (melting point) than a perfect match will. With the LightCycler, PCR amplification and fluorescence detection can be done in the same tube within 60 min. The LightCycler is a convenient and economic platform for parallel detection of the three mutations *(61)*. Of note, mutations within the Factor V gene can lead to a false-positive result.

4.6. DNA CHIPS With recent advances in gene chip technology, automated platforms are now available in which PCR products are hybridized to an array of probes to screen for multiple mutations. A DNA chip generally consists of DNA sequences representative of a genome of interest bound to a solid support. The target DNA to be tested is fluorescently labeled and, hybridized to the array, and the differential fluorescence is detected by laser scanning with confocal microscopy *(65)*.

The Affymetrix chip is one type of DNA microchip in which short oligonucleotide probes with known nucleotide sequences are bound to a solid support surface *(66)*. Affymetrix designed chips that contain oligonucleotide probes with many known SNPs, so-called "SniP chips," have been developed. These DNA microchips might provide a profile array that can be used to analyze gene expression levels in several different coagulation disorders *(65,66)*.

Although global gene expression through the use of DNA microchips such as the Affymetrix GeneChip can provide valuable information, use of a small panel of genes that can provide rapid diagnosis of specific targets of interests have been emerging as a diagnostic tool. Nanogen Inc. has developed a DNA

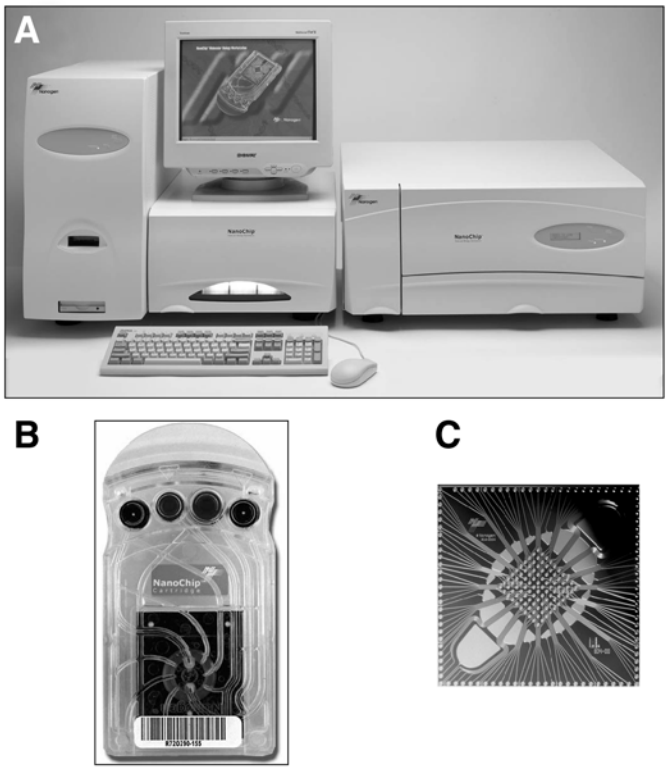


Fig. 4. The Nanogen Nanochip workstation **(A)**, microarray cartridge **(B)** and electronic pads **(C)**. For more information, see http://www.nanogen.com.

microchip wherein small subsets of target DNA can be transferred, denatured, and hybridized with the use of electric fields *(62)*.

In contrast to DNA chips with a huge array of probes, the NanoChip provides a platform in which a small percentage of expressed genes from different patients can be analyzed. The NanoChip® (Fig. 4) is based on an electronic microarray in which voltage and current are applied for electronic field control. By confining hybridization to a specific site on a microarray, multiple genetic analysis can be done on a single microarray by distributing different assays to specific subsets of array sites *(62)*. SNPs for different genes from different samples can all be detected in the single platform provided by the NanoChip. These small electronic chips have become recently available for FVL, Prothrombin G20210A, and MTHFR mutation analysis *(67)*.

In the NanoChip system, biotinylated oligonucleotide capture probes are electronically located on specific sites on the NanoChip cartridge. After multiplexed PCR for FVL and G20210A, the amplified PCR product is transferred to a well plate and addressed electronically to pads on the Nanochip cartridge for hybridization to the capture probes. Results are read with fluorescently labeled wild and mutant reporter oligonucleotides; the intensity of the fluorescent signal correlates with the amount of target hybridized on the microarray *(62)*. The use of an electric field permits improved hybridization.

Advantages of the NanoChip over the usual RFLP and LightCycler assay is very high throughput, automation, and a comprehensive results display *(67)*. However, compared to the LightCycler, the NanoChip assay has a longer turnaround time. The NanoChip is currently recommended for laboratories with medium to large test volume *(67)*.

4.7. NON-PCR-BASED HYBRIDIZATION ASSAY Direct hybridization methods without prior amplification with PCR are also available for DNA analysis. A signal amplification system known as the Invader assay (Third Wave Technologies, Madison, WI) can detect mutations or SNPs using a combination of hybridization technique with enzyme recognition *(63)*. The assay is based on the formation of a three-stranded product comprised of a downstream signal probe, the target DNA template with normal or mutant nucleotide sequence, and the upstream Invader probe; this product is then cleaved by an enzyme, resulting in a measurable product *(63)*.

The enzyme used in the Invader assay is referred to as Cleavase. This enzyme cleaves DNA molecules within a given temperature range and at specific recognizable structures generated by the hybridization of the two oligonucleotide probes to the target DNA. Detection of the Invader reaction products can be achieved by gel electrophoresis, or enzyme-linked immunoassay *(63)*.

The ability of the Invader assay to detect a mutation is based on the substrate specificity of the cleavage enzyme used. Probe design is simpler because specificity does not rely entirely on the hybridization step but also on enzyme substrate recognition. Sample processing is simplified because the need for target amplification is eliminated. However, methods based on probe hybridization alone, such as the Invader assay, are generally considered less sensitive than PCR *(56)*.

4.8. COMPARING MOLECULAR VS NONMOLECULAR TESTING FOR THE FACTOR V LEIDEN MUTATION Two different diagnostic approaches—a clot-based assay and a molecular assay—can be used for detection of Factor V Leiden mutation *(68)*. The functional clot-based assays for evaluation of the phenotypic abnormality associated with the FV Leiden genetic defect (APC resistance) have been available as first- and second-generation assays.

The first-generation tests are not as sensitive as the second-generation assays or DNA-based analysis; however, APC resistance caused by mutations other than the FVL mutation can be detected by this first-generation assay. This has raised the question of the significance of cases being missed by not performing first-generation clot-based assays. The clinical importance of detecting these non-FVL cases is uncertain. A recent study concluded that measurement of APC resistance *without* dilution in FV-deficient plasma (a typical first-generation clot-based assay) was necessary to evaluate for potentially important risk factors associated with thrombosis other than FVL *(69)*.

On the other hand, Anticoagulation treatment, coagulation inhibitors, and acute thrombosis all can variably interfere with the first-generation clot-based assays for APC resistance; in contrast, the current second-generation APC assays and molecular-based testing can both be performed in these circumstances. In addition, the second-generation assay and molecular genetic testing can distinguish between heterozygotes and homozygotes *(68,70)*. Distinguishing between the heterozygote and homozygote state is important because of the risk associated with each FVL carrier state: an overall 3- to 7-fold

increased risk of venous thrombosis for heterozygotes and a 50- to 100-fold increased risk for homozygotes *(71)*.

Although a FVL functional clot-based assay is proposed by some as an appropriate starting point for the evaluation of APC resistance phenotype, the best approach to FVL analysis remains controversial *(68)*. At our institution, testing for the FVL is done by DNA-based testing without performing a functional first or second-generation clot-based assay.

4.9. ADVANTAGES/DISADVANTAGES OF MOLECULAR TESTING The major advantages of molecular genetic testing are increased specificity and sensitivity. To some degree, the challenges posed on conventional clot-based coagulation assays given the dynamic nature of the coagulation process is circumvented with direct genetic analysis. DNA-based assays can be performed while the patient is on anticoagulation treatment, in the presence of specific or nonspecific inhibitors, including lupus anticoagulant, and in the setting of a recent thrombosis where expected consumption of coagulation proteins occurs regardless of the presence of a hereditary defect. In addition, as mentioned earlier, molecular testing can distinguish between heterozygotes and homozygotes.

Limitations of molecular testing include those generally known for all PCR-based assays, including the need to know the specific sequence of the DNA of interest, PCR contamination, and specificity resulting from nonspecific DNA products *(56)*. In theory, all mutation detection methods not based on sequencing could give misleading results if there are other base exchanges in the vicinity of the mutation of interest. False positives caused by mutations or polymorphisms that alter a different basepair at the same restriction site when using RFLP have been described *(56)*. False-negative PCR assays can occur in patients with the APC resistance phenotype because of mutations other than Factor V Leiden, albeit the significance of these mutations is controversial.

As mentioned previously, molecular genetic testing for antithrombin, protein C, and protein S deficiency is currently impractical given the inability to screen for all the mutations associated with these defects.

5. CLINICAL APPLICATIONS OF MOLECULAR DIAGNOSTICS IN THE COAGULATION LABORATORY

The increased demand for thrombophilia testing is explained in part by the relatively common occurrence of venous thromboembolism (VTE): VTE is the third most common cardiovascular disease in the United States and accounts for 100,000 deaths annually *(72)*. Subsets of these patients have a hereditary or acquired predisposition for thrombosis referred to generally as thrombophilia. Thrombophilia could be an important cause of unprovoked idiopathic or recurrent VTE, thrombosis in patients <50 yr of age, thrombosis at unusual sites, family history of VTE, and complications of pregnancy. Up to 70% of patients with thrombophilia might have one or more of the major inherited coagulation defects. In this context, molecular genetic testing for hereditary defects associated with thrombophilia has become readily available.

FVL mutation is the most common hereditary thrombophilia. In patients with recurrent venous thromboembolism, 40–50% are heterozygous for the Factor V Leiden *(15,68)*. In a prospective study of a large cohort of apparently healthy men, the FVL mutation was associated with an increased risk of venous thrombosis, whereas no increased risk was seen for myocardial infarction or stroke *(73)*. Of note, although testing for the FVL mutation might determine the risk of thrombosis in a patient, this analysis, in turn, is limited because of the low predictive value of the test *(74)*. Furthermore, although the homozygous state for FVL might be associated with an 80-fold increased risk for thrombosis *(15)*, this in itself does not represent a disease state *(74)*. In contrast, the homozygous states for protein C and S deficiencies are usually associated with a severe clinical condition such as neonatal purpura fulminans *(20)*. Compared to antithrombin, protein C, and protein S deficiencies, FVL has a higher prevalence among the white population, but a smaller risk associated with thrombosis *(75)*.

A positive genetic test has variable implications depending on the selected population. Screening for the FVL mutation could have important clinical benefits in selected patients, such as patients with central venous catheter-associated thrombosis *(76)* or in renal transplant recipients *(77)*. On the other hand, the benefits, if any, of screening for the FVL mutation in other populations is not known. Although the combined effect of FVL and use of oral contraceptives might have a multiplicative effect on the risk of VTE *(78)*, there is no consensus on routine FVL screening for all women considering use of oral contraceptives *(2)*. In a recent consensus conference organized by the College of American Pathologists, screening for the FVL in women before use of oral contraceptives was not generally recommended *(68)*.

In some patients, combined genetic defects might be present. A positive FVL test along with another hereditary defect(s) appears to compound the risk of thrombosis *(79,80)*. Individuals with coexistent FVL and prothrombin mutations *(79)* and FVL and protein C deficiency *(80)* are at an even higher risk of thrombosis than with a single gene defect. As a result, testing for the FVL and prothrombin 20210 mutations along with antithrombin, protein C, and protein S deficiencies are all recommended for hereditary thrombophilia work-up, because a combined defect might affect clinical decision-making.

The presence of the FVL mutation could also indicate that other family members might be at risk of developing venous thrombosis *(68)*; therefore, FVL testing has been extended to family members of patients who experienced VTE and tested positive for the FVL. Although testing for the FVL mutation in this population has the potential to identify asymptomatic carriers who might benefit from thromboprophylaxis during periods of high risk, the reportedly low rate of VTE among family members with the FVL mutation did not justify the use of prophylactic anticoagulation *(81)*.

The prothrombin gene mutation G20210A is the second most common hereditary thrombophilic defect, with a prevalence in the general population as high as 1–4% *(42,82)*. The prothrombin mutation is a well-established risk factor for venous thrombosis *(83)*; its role in arterial thrombosis remains controversial *(84)*. Similar to the FVL, the prothrombin mutation has a higher prevalence than antithrombin, protein C, and protein S deficiencies, but a lower risk associated with thrombosis.

Table 2
College of American Pathologists Consensus Recommendations for Factor V Leiden and Prothrombin 20210A Testing

History of recurrent VTE
VTE before the age of 50 yr
Unprovoked VTE at any age
VTE at unusual sites
VTE patients at any age with a positive family history (first-degree relative with VTE before the age of 50)
VTE secondary to pregnancy, oral contraceptives, or hormone replacement therapy

Routine screening of the general population or as an initial screening test during pregnancy or before use of oral contraceptives is not recommended *(64)*. However, in select patient populations, testing for the prothrombin mutation is advised *(64)*.

The third most commonly performed genetic test for thrombophilia is the methylenetetrahydrofolate reductase (MTHFR) mutation analysis. Based on prospective studies, there appears to be a weak association between hyperhomocysteinemia and arterial thrombosis; the association with venous thrombosis is controversial *(52)*. Whereas homozygosity for the MTHFR mutation is associated with elevated levels of homocysteine, the mutation itself is not associated with cardiovascular disease *(85)*. Homocysteine levels are recommended for patients with documented atherosclerotic disease; however, given the lack of evidence of homozygosity being an independent risk factor for arterial and venous thrombosis, genotyping for either the C677T or A1298C mutations are not generally recommended *(52)*.

Molecular diagnostics in the area of coagulation is not exempt from the ongoing controversy and debate surrounding testing in thrombophilic disorders. The impact of a positive PCR test for Factor V Leiden mutation on therapeutic decision-making (life-long anticoagulation?) as well as the potential financial consequences (reduced insurability) are reasons for much concern surrounding the appropriate use of these tests. Based mostly on retrospective studies or multiple anecdotal studies, the College of American Pathologists consensus conference recommended molecular testing for FVL and prothrombin in selected patients (Table 2).

6. CONCLUSION

Much change is anticipated in the area of coagulation testing. The "uncoding" of the human genome will continue to provide an important amount of information that will likely affect the diagnostic approach to various diseases, including coagulation disorders. Currently, most of the applications of molecular diagnostics in coagulation have been confined to the analysis of one gene and its corresponding mutation or polymorphism. New technology based on DNA microchips, in which different genes could be looked at simultaneously, might provide a more complete genetic profile of coagulation disorders.

The role of molecular diagnostics for the hemophilias might extend beyond carrier detection and prenatal diagnosis; the identification of specific subgroups of hemophiliac patients with distinct clinical syndromes might be clinically relevant. Likewise, the use of molecular diagnostics in Von Willebrand disease is currently limited and remains fertile ground for potential applicability of routine molecular genetic testing for the diagnosis and subclassification of this disorder.

Molecular diagnostics in thrombophilia work-up will continue to expand. A gene panel in which patterns of multiple genes are evaluated could provide a more complete assessment of thrombotic risk in a given patient compared to analysis of one or two individual gene defects. Multigene platforms such as DNA microchips may become available for evaluation of thrombophilia.

Furthermore, many genes have shown a highly heterogeneous mix of mutations for which testing by the traditional methods is difficult, expensive, and sometimes inconclusive. Protein C deficiency is known to be associated with more than 100 mutations; it is conceivable that in a gene chip all known mutations could be tested. Identifying a specific type of mutation might correlate with the severity of the disease. Likewise, determining which genetic risk factor an individual patient has inherited might also be clinically relevant.

However, whether analysis of an entire set of expressed genes offers another diagnostic perspective for disorders of hemostasis and thrombosis and allows for a more complete classification of coagulation disorders remains to be determined. Meanwhile, automation of DNA technology along with combination of multiple genetic markers in a single assay will continue to reduce the costs of molecular tests and make a strong case for PCR-based and non-PCR-based molecular assays to replace many traditional coagulation tests.

In light of the increasing number of genetic mutations and polymorphisms being described in coagulation disorders, the challenge for the coagulation and molecular diagnostics laboratory together will be to determine what test(s) to offer and how to best test for these defects in a manner relevant to patient care.

ACKNOWLEDGMENT

The author thanks Dr. Kojo S.J. Elenitoba-Johnson (Department of Pathology, University of Utah, Salt Lake City) for critical review of the manuscript.

REFERENCES

1. Olson, J. D. College of American Pathologists Consensus Conference XXXVI: Diagnostic Issues in Thrombophilia. *Arch. Pathol. Lab. Med.* **126:**1277–1279, 2002.
2. Khoury, M. J., McCabe L. L., and McCabe E. R. B. Genomic medicine: population screening in the age of genomic medicine. *N. Engl. J. Med.* **348(1):**50–58, 2003.
3. Mann, K. G. Biochemistry and physiology of blood coagulation. *Thromb. Haemost.* **82:**165–174,1999.
4. Brandt, J. T. Overview of hemostasis, in *Clinical Laboratory Medicine*, McClatchey, K. D., ed., Lippincott Williams & Wilkins, Philadlephia, PA, 2002, pp. 988–1009.
5. de Groot, P. G. The role of von Willebrand factor in platelet function. *Sem in Thromb. Hemost.* **28:**133–138, 2002.
6. Morrissey, J. H. Tissue factor and factor VII initiation of coagulation, in *Hemostasis and Thrombosis,* Colman, R. W., Hirsh, J., Marder, V. J., Clowes, A. W.,and George, J. N. ed., Lippincott Williams and Wilkins, Philadelphia, PA, 2001, pp. 89–101.
7. Esmon, C. T. The roles of protein C and thrombomodulin in the regulation of blood coagulation. *J. Biol. Chem.* **264:**4743–4746, 1989.

8. Collen, D. The plasminogen (fibrinolytic) system. *Thromb. Haemost.* **82:**259-270, 1999.
9. Tefferi, A., Wieben, E., Dewald, G. W., Whiteman, D. A. H., Bernard, M. E., and Spelsberg, T. C. Primer on medical genomics: Part II: Background principles and methods on molecular genetics. *Mayo Clin. Proc.* **77:**785–808, 2002.
10. Rennert, H. and Leonard, D. G. B. Basic principles of molecular biology, in *Diagnostic Molecular Pathology, W. B.* Saunders, Philadelphia, PA, 2003, pp. 25–51.
11. Reitsma, P. H. Genetic Principles underlying disorders of procoagulant and anticoagulant proteins, in *Hemostasis and Thrombosis,* Colman, R. W., Hirsh, J., Marder, V. J., Clowes, A. W., and George, J. N. ed., Lippincott Williams and Wilkins, Philadelphia, PA, 2001, pp. 59–87.
12. Lane, D. A. and Grant, P. J. Role of hemostatic gene polymorphisms in venous and arterial thrombotic disease. *Blood* **95:**1517–1532, 2000.
13. Kottke-Marchant, K. Genetic polymorphisms associated with venous and arterial thrombosis. *Arch. Pathol. Lab. Med.* **126:**295–304, 2002.
14. Franco, R. F. Gene polymorphism of the haemostatic system and the risk of arterial thrombotic disease. *Br. J. Haematol.* **115:**491–506, 2001.
15. Bertina, R. M. Molecular risk factors for thrombosis. *Thromb. Haemost.* **82:**601–609, 1999.
16. Humphries, S. E., Panahloo, A., Montgomery, H.E., Green, F., and Yudkin, J. Gene–environment interaction in the determination of levels of haemostatic levels involved in thrombosis and fibrinolysis. *Thromb. Hemost.* **78:**457–461, 1997.
17. Seligsohn, U. and Zivelin, A. Thrombophilia as a multigenic disorder. *Thromb. Haemost.* **98:**297–301, 1997
18. Rosendaal, F. R. Venous thrombosis: a multicausal disease. *Lancet* **353:**1167,1999.
19. Ridker, P. M., Miletich, J. P., Hennekens, C. H., and Buring, J. E. Ethnic distribution of factor V Leiden in 4047 men and women. Implications for venous thromboembolism screening. *JAMA* **277:**1305–1307, 1997.
20. Marlar, R. A. and Neumann, A. Neonatal purpura fulminans due to homozygous protein C or protein S deficiencies. *Semin. Thromb. Haemost.* **16:**299-310, 1990.
21. Tuddenham, E. G. D., Schwaab, R., Seehafer, J., et al. Haemophilia A : database of nucleotide substitutions, deletions, insertions, and rearrangements of the factor VIII gene, second edition. *Nucleic Acids Res.* **22:**4861–4868, 1994.
22. Kazazian, H. H., Wong, C., Youssoufian, H., Scott, A. F., Phillips, D. G., and Antonarakis, S. E. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* **332:**164–166, 1998.
23. Fijnvandraat, K., Turenhout, E. A., van den Brink, E. N., et al. The missense mutation Arg503→Cys is related to antibody formation in a patient with mild hemophilia A. *Blood* **89:** 4371–4377, 1997.
24. Gianelli, F., Green P. M., Sommer, S. S., et al. Haemophilia B: database of point mutations and short addition and deletions-eighth edition. *Nucleic Acid Res.* **26:**265–268, 1998.
25. Briet, E., Bertina, R. M., van Tilberg, N. H., and Veltkamp, J. J., Haemophilia B Leiden: a sex linked hereditary disorder that improves after puberty. *N. Engl. J. Med.* **306:**788–790, 1982.
26. Oldenberg, J., Quenzel, E. M., Harbrecht, U., et al. Missense mutations at ALA-10 in the factor IX propeptide: an insignificant variant in normal life but a decisive cause of bleeding during oral anticoagulant therapy. *Br. J. Haematol.* **98:**240-244, 1997.
27. Nichols W. C., and Ginsburg, D. Von Willebrand disease. *Medicine* **76:**1–20, 1997.
28. Schneppenheim, R., Budde, U., and Ruggeri, Z. M. A molecular approach to the classification of von Willebrand disease. *Baillieres Clin. Haematol.* **14:**281–298, 2001.
29. Cooper, D. N., Millar, D. S., Wacey, A., et al. Inherited factor VII deficiency: molecular genetics and pathophysiology. *Thromb. Hemost.* **78:**151–160, 1997.
30. Cooper, D. N., Millar, D. S., Wacey, A., et al. Inherited factor X deficiency: molecular genetics and pathophysiology. *Thromb. Hemost.* **78:**161–172, 1997.
31. Asakai, R., Chung, D. W., Davie, E. W., and Seligsohn, U. Factor XI deficiency in Ashkenasi Jews in Israel. *N. Engl. J. Med.* **326:** 153–158, 1996.
32. Schloesser, M., Zeerleder, S., Lutze, G., et al. Mutations in the human factor XII gene. *Blood* **90:**3967–3977, 1997.
33. Anwar, R. and Miloszewski, K. J. Factor XIII deficiency. *Br. J. Haematol.* **107:**468–484, 1999.
34. Cunningham, M. T., Brandt, J. T., Laposata, M., and Olson, J. D. Laboratory diagnosis of dysfibrinogenemia. *Arch. Pathol. Lab. Med.* **126:**499–505, 2002.
35. Dahlbach, B., Carlsson, M., and Svensson, P. J., Familial thrombophilia due to a previously unrecognized mechanism characterized by poor anticoagulant response to activated protein C: prediction of a cofactor to activated protein C. *Proc. Natl. Acad. Sci. USA* **90:**1004–1008, 1993.
36. Bertina, R. M., Koeleman, B. P., Koster, T., et al. Mutation in blood coagulation factor V associated with resistance of activated protein C. *Nature* **369:**64–67, 1994.
37. Williamson, D., Brown, K., Luddington R., Baglin, C., and Baglin, T. Factor V Cambridge, a new mutation (arg306-Thro) associated with resistance of activated Protein C. *Blood* **91:**1140–1144, 1998.
38. Chan, W. P., Lee, C. K., Kwong, Y. L., et al. A novel mutation of Arg 306 of factor V gene in Hong Kong Chinese. *Blood* **91:**1135–1139, 1998.
39. de Visser, M. C. H., Rosendaal, F. R., and Bertina, R. M. A reduced sensitivity for activated protein C in the absence of factor V Leiden increases the risk of venous thrombosis. *Blood* **93:**1271, 1999.
40. Bernardi, F., Faioni, E. M., Castoldi, E., Castaman, G., Sacchi, E., and Mannucci, P. M. A factor V genetic component differing from factor V R506Q contributes to the activated protein C resistance phenotype. *Blood* **90:**1552–1557, 1997.
41. Alhenc-Gelas, M., Nicaud, V., Gandrille, S., et al. The factor V gene A4070G mutation and the risk of venous thrombosis. *Thromb. Haemost.* **81:**193, 1999.
42. Poort, S. R., Rosendaal, F. R., Reitsma, P. H., and Bertina, R. M. A common genetic variation in the 3-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. *Blood.* **88:**3698–3703, 1996.
43. Gehring, N. H., Frede, U., New-Yilik, G., et al. Increased efficiency of mRNA 3end formation: a new genetic mechanism contributing to hereditary thrombophilia. *Nat. Genet.* **28:**389–392, 2001.
44. Egeberg, O. Inherited antithrombin deficiency causing thrombophilia. *Thromb. Diath. Haemorrh.* **12:**516–530, 1965.
45. Finazzi, G., Caccia, R., and Barbui, T. Different prevalences of thromboembolism in the subtypes of congenital antithrombin III deficiency: review of 404 cases. *Thromb. Haemost.* **58:**1094, 1987.
46. Reitsma, P. H., Bernardi, F., Doig, R. G., et al. Protein C deficiency: a database of mutations, 1995 update. On behalf of the subcommittee on Plasma Coagulation Inhibitors of the Scientific and Standardization Committee of the ISTH (1991). *Thromb. Hemost.* **73:**876–889, 1995.
47. Gandrille, S., Borgel, D., Ireland, H., et al. Protein S deficiency: a database of mutations. For the Plasma Coagulation Inhibitors Subcommittee of the Scientific and Standardization Committee of the International Society of Thrombosis and Hemostasis. *Thromb. Haemost.* **77:**1201–1214, 1997.
48. Ploos van Amstel, J. K., van der Zanden, A. L., Bakker, E., Reitsma, P. H., and Bertina, R. M. Two genes homologues with human protein S cDNA are located on chromosome 3. *Thromb. Haemost.* **58:**982–987, 1987.
49. Ohlin, A. K., Norlund, L., and Marlar, R. A. Thrombomoduliin gene variations and thromboembolic disease. *Thromb. Haemost.* **78:**396–400, 1997.
50. Van der Velden, P. A., Krommenhoek-Van Es, T., Alaart , C. F., Bertina, R. M., and Reitsma, P. H. A frequent thrombomodulin amino acid dimorphism is not associated with thrombophilia. *Thromb. Haemost.* **65:**511–513, 1991.
51. Rosendaal, F. R., High levels of VIII and venous thrombosis. *Thromb. Haemost.* **83:**1–2, 2000.

52. Nigel, S. K. and McGlennen, R. C. Hyperhomocysteinemia and thrombophilia. *Arch. Pathol. Lab. Med.* **126:**1367–1375, 2002.
53. Froost, P., Blom, H. J., Milos, R., et al. A candidate genetic risk factor for vascular disease. A common mutation in methylenetetrahydrofolate reductase. *Nat Genet.* **10:**111–113, 1995.
54. Hanson, N. Q., Aras, O., Yang, F., and Tsai, M. Y. C677T and A1298C polymorphisms of the methylenetetrahydofolate reductase gene: incidence and effect of combined genotypes on plasma fasting and post methionine load homocysteine in vascular disease. *Clin. Chem.* 47:661–666, 2001.
55. Lutz, C. T., Foster, P. A., Noll, W. W., et al. Multicenter evaluation of PCR methods for the detection of Factor V Leiden (R506Q) genotypes. *Clini Chem.* **44:**1356–1357, 1998.
56. Rennert, H. and Leonard, D. G. B. Molecular methods in the diagnostic laboratory, in *Diagnostic Molecular Pathology,* Leonard, D. G. B., ed., W. B. Saunders, Philadelphia, PA, 2003, pp. 25–51.
57. Gomez, E., van der Poel, S., Jansen, J. H., van der Reijden, B. A., and Lowenberg, B. Rapid simultaeous screening of Factor V Leiden and G20210A prothrombin variant by multiplex polymerase chain reaction on whole blood. *Blood* **91:**2208–2209, 1998.
58. Hezard, N., Cornillet-Lefebvre, P., Gillot, L., Potron, G., and Nguyen, P. Multiplex ASA PCR for a simultaneous detection of factor V Leiden gene, G-A 20210 prothrombin gene and C-T MTHFR gene mutations. *Thromb. Haemost.* **79:**1054–1055, 1998.
59. Endler, G., Kyrle, P. A., Eichinger, S., Exner, M., and Mannhatler, C. Multiplex mutagenically separated PCR: simultaneous single tube detection of the factor V R506Q (G1691A), the prothrombin G20210A, and the methylenetetrahydrofolate reductaseA223V (C677T) variants. *Clin. Chem.* **47:**333–335, 2001.
60. Kirschbaum, N. E. and Foster, P. A. The polymerase chain reaction with sequence specific primers for the detection of the factor V mutation associated with activated protein C resistance. *Thromb. Haemost*. **74:**874–878, 1995.
61. van Ahsen, N., Shutz, E., Armstrong, V.W. and Oellerich, M. Rapid detection of prothrombotic mutations of prothrombin (G20210A), factor V (G1691A), and methylenetetrahydrofolate reductase (c677T) by real-time fluorescence PCR with the LightCycler. *Clin. Chem.* **45:**694–696, 1999.
62. Weidenhammer, E. M., Kahl, B. F., Wang, L., et al. Multiplexed, targeted gene expression profiling and genetic analysis on electronic microarrays. *Clin. Chem.* **48:**1873–1882, 2002.
63. Kwiatkowski, R. W., Lyamichev, V., de Arruda, M., and Neri, B. Clinical, genetic, and pharmacogenetic applications of the Invader assay. *Mol. Diagn.* **4:**354–364, 1999.
64. McGlennen, R. C. and Key, N. S. Clinical and laboratory management of the prothrombin G20210 mutation. *Arch. Pathol. Lab. Med.* **26:**1319–1325, 2002.
65. Miesfeld, R. L. Contemporary applied molecular genetics, in *Applied Molecular Genetics,* Miesfeld, R. L., ed., Wiley–Liss, New York, 1999, pp. 237–265.
66. Wallace, R. W. DNA on a chip: serving up the genome for diagnostics and research. *Mole. Med. Today* **3:**384–389, 1997.
67. Schrijver, I., Lay, M. J., and Zehnder, M.D. Diagnostic single nucleotide polymorphism analysis of Factor V Leiden and prothrombin 20210G>A. *Am. J. Clin. Pathol.* **119:**490–496, 2003.
68. Press, R., Bauer, K., Kujovich, J. L., and Heit, J. A. Clinical utility of Factor V Leiden R506Q testing for the diagnosis and management of thromboembolic disorders. *Arch. Pathol. Lab. Med.* **126:**1304–1318, 2002.
69. Graf, L., Welsh, C. H., Qamar, Z., Marlar, R. A. Activated Protein C Resistance Assay detects thrombotic risk factors other than factor V Leiden. *Am. J. Clin. Pathol.* **119:**522–60.
70. Legnani, C. C., Palareti, G., Biagi, R., et al. Activated protein C resistance: a comparison between two clotting assays and their relationship to the presence of factor V Leiden mutation. *Br. J. Hematol.* **93:**694–699, 1996.
71. Dahlbach, B. New molecular insights intp the genetics of thrombophilia: resistance to activated protein C caused by Arg506 to Gln mutation in factor V as a pathogenic risk factor for venous thrombosis. *Thromb. Haemost.* **74:**139–148.
72. Heit, J. A., Silverstein, M. D., Mohr, D. N., et al. The epidemiology of venous thormobemboliism in the community *Thromb. Haemost.* **86:**452–463, 2001.
73. Ridker, P. M., Hennekens, C. H., Lindpaintner,K., Stampfer, M. J., Eisenberg, P. R., and Miletich, J. P. Mutation in the gene coding for coagulation factor V and the risk of myocardial infarction, stroke, and venous thrombosis in apparently healthy men. *N. Engl. J. Med.* **332:**912–917, 1995.
74. Burke, W. Genomic medicine: genetic testing. *N. Engl. J. Med.* **347:** 1867–1875, 2002.
75. Martinelli, I., Mannucci, P. M., de Stefano, V., et al. Different risks of thrombosis in four coagulation defects associated with inherited thrombophilia; a study of 150 families. *Blood* **92:**2353, 1998.
76. Fijnheer, R., Paijmans, B., Verdonck, L. F., Nieuwenhuis, H. K., Roest, M., and Dekker, A. W. Factor V Leiden in central venous catheter-a ssociated thrombosis. *Br. J. Haematol.* **118:**267–270, 2002.
77. Friedman, G. S., Meier-Kriesche, H. U., Kaplan, B., et al. Hypercoagulable states in renal transplant candidates: impact of anticoagulation upon incidence of renal allograft thrombosis. *Transplantation* **27:**1073–1078, 2001.
78. Vaandenbrouke, J. P., Koster, T., Briet, E., Reitsma, P. H., Bertina, R. M., Rosendaal, F. R., Increased risk of venous thrombosis in oral contraceptives users who are carriers of factor V Leiden mutation. *Lancet* **344:**1453–1457, 1994.
79. Emmerich, J., Rosendaal, F. R., Catanneo, M. et al. Combined effect of factor F Leiden and prothrombin 20210A on the risk of venous thromboembolism: pooled analysis of 8 case control studies including 2310 cases and 3204 controls: study group for Pooled Analysis in Venous Thromboembolism. *Thromb. Haemost.* **86:**809–816, 2001.
80. Koeleman, B. P. C., Reitsma, P. H., Allart, C. F., and Bertina, R. M. Activated protein C resistance as an additional risk factor for thrombosis in protein C deficient families. *Blood* **84:**1031–1035, 1994.
81. Simioni, P., Tormene, D., Prandoni, P., et al. Incidence of venous thromboembolism in asymptomatic family members who are carriers of factor V Leiden: a prospective cohort study. *Blood* **99:**1938–1942, 2002.
82. Rosendaal, F. R., Doggern, C. J., Zivelin, A., et al. Geographic distribution of the 20210 G to A prothrombin variant. *Thromb. Haemost.* **79:**706–708, 1998.
83. Girolami, A., Simioni, P., Scarano, L., and Carraro, G. Prothrombin and the prothrombin 20210 G to A polymorphism: their relationship with hypercoagulability and thrombosis. *Blood Rev.* **13:**205–210, 1999.
84. Franco, R. F. Gene polymorphism of the haemostatic system and the risk of arterial thrombotic disease. *Br. J. Haematol.* **115:**491–506, 2001.
85. Brattstrom, L., Wilcken, D. E., Ohrvik, J., and Brudin, L. Common methylenetetrahydrofolate reductase gene mutation leads to hyperhomocysteinemia but not to vascular disease: the result of a meta-analysis. *Circulation* **98:**2520–2526, 1998.

27 Cystic Fibrosis

Eugene H. Lewis III, Myra J. Lewis, Jean A. Amos, and Gregory J. Tsongalis

1. INTRODUCTION

Cystic fibrosis (CF), a clinically heterogeneous disease, is the first genetic disease for which adult population screening has been initiated in the United States. Since the discovery of the *CFTR* gene in 1989, much has been learned about the pathophysiology and molecular genetics of this disorder. This review includes an overview of the genetics of CF, a discussion of pathophysiology, and clinical and anatomic pathology and concludes with a review of molecular diagnostics.

2. GENETICS

Cystic fibrosis is the most common lethal autosomal-recessive disease in individuals of European decent with a prevalence of 1 in 2500 to 3300 live births. Although most common in the Caucasian population, it has become apparent that members of other racial and ethnic backgrounds are also at risk *(1)*. Approximately 30,000 children and adults in the United States are affected, and approx 850 individuals are newly diagnosed annually with the majority of those less than 1 yr of age.

The cystic fibrosis transmembrane conductance regulator (*CFTR*) gene, the gene responsible for CF, spans approx 230 kb on chromosome 7q. This gene consists of 27 coding exons that result in a 6.5-kb mRNA product and a CFTR membrane glycoprotein of 1480 amino acids with a mass of approx 170,000 Da *(2–4)*. CFTR functions as a cAMP- regulated chloride channel in the apical membrane of epithelial cells *(5)*.

To date, over 1000 unique mutations in the *CFTR* gene have been described *(6)*. The most common mutation is the deletion of phenylalanine at position 508 (ΔF508) and affects 70% of patients worldwide. Approximately half of CF patients are ΔF508 homozygotes. However, the majority of mutations are private, occurring only in single families, or are rare. G542X is the most frequent Caucasian mutation after ΔF508, occurring at a frequency of 2%. The allelic frequency of CFTR mutations varies by ethnic group, where, for example, ΔF508 is only present in 30% of the Ashkenazi Jewish population.

The mutations in the *CFTR* gene are grouped into six classes. These include class I (defective protein synthesis, where there is no CFTR protein at the apical membrane), class II (abnormal/ defective processing and trafficking, where there is no CFTR protein at the apical membrane), class III (defective regulation, where there is a normal amount of nonfunctional CFTR at the apical membrane), class IV (decreased conductance, where there is a normal amount of CFTR with some residual function at the apical membrane), class V (reduced or defective synthesis/trafficking, where there is a decreased amount of functional CFTR at the apical membrane), and class VI (decreased stability, where there is a functional but unstable CFTR at the apical membrane) *(7,8)*. Of the CFTR mutations, classes I–III are the most common and are associated with pancreatic insufficiency *(9)*. The most common mutation worldwide is the ΔF508, class II, with varying frequency among ethnic groups *(10)*.

Mutations in the *CFTR* gene can lead to an abnormal protein that causes defective electrolyte transport and defective chloride ion transport in apical membrane epithelial cells affecting the respiratory tract, pancreas, intestine, male genital tract, hepatobiliary system, and the exocrine system, resulting in complex multisystem disease. CFTR is a member of an ATP-binding cassette family with diverse functions such as ATP-dependent transmembrane pumping of large molecules, regulation of other membrane transporters, and ion conductance. The loss of CFTR-mediated anion conductance explains a variety of CF symptoms, including elevated sweat chloride because of defective salt absorption by the sweat ducts and meconium ileus because of defective fluid secretion by intestinal crypt cells *(11)*. The malfunction of CFTR as a regulator of amiloride-sensitive epithelia Na^+ channel leads to increased Na^+ conductance in CF airways, which drives increased absorption of Cl^- and water. Most symptoms of CF, such as meconium ileus, loss of pancreatic function, degeneration of the vas deferens, thickened cervical mucus, and failure of adrenergically mediated sweating are the result of the role CFTR plays in Cl-driven fluid secretion.

3. CLINICAL TESTING

A diagnosis of CF in a symptomatic or at-risk patient is suggested by clinical presentation and confirmed by a sweat test.

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

Table 1
Recommended Mutation Panel for Cystic Fibrosis Carrier Screening

ΔF508	*ΔI507*	*G542X*	*G551D*	*W1282X*	*N1303K*
R553X	621 + 1G > T	R117H	1717–1G > A	A455E	R560T
R1162X	G85E	R334W	R347P	711 +⊇1G > T	1898 + 1G > A
2184delA	1078delT	3849 + 10kbC > T	2789 + 5G > A	3659delC	I148T
3120 + 1G>A					
I506V[a]					
I507V[a]					
F508C[a]					
5T/7T/9T[a]					

[a]Reflex tests.

In the presence of clinical symptoms (e.g., recurrent respiratory infections), a sweat chloride above 60 mmol/L is diagnostic for CF. Although the results of this test are valid in a newborn as young as 24 h, collecting a sufficient sweat sample from a baby younger than 3 or 4 wk old is difficult. The sweat test can also confirm the diagnosis in older children and adults but is not useful for carrier detection. *CFTR* mutation analysis is subsequently performed to identify the molecular genetic abnormality and to confirm sweat testing results.

4. PATHOPHYSIOLOGY

CFTR is an anion channel that functions in the regulation of ion transport. It plays multiple roles in fluid and electrolyte transport, including salt absorption, fluid absorption, and anion-mediated fluid secretion *(11)*. Defects in this protein, as described earlier, lead to CF, the morbidity of which is initiated by a breach in host defenses and propagated by an inability to clear the resultant infections *(12)*. Because inflammatory exacerbations precipitate irreversible lung damage, the innate immune system plays an important role in the pathogenesis of CF.

Respiratory epithelial cells containing the *CFTR* gene also provide a crucial environmental interface for a variety of inhaled insults. The local mucosal mechanism of defense involves mucociliary clearance that relies upon the presence and constituents of airway surface liquid (ASL). The high salt in the ASL of CF patients is found to interfere with the natural antibiotics present in ASL, such as defensins and lysozyme *(13)*. Bales et al. categorizes the role of CFTR in the pathogenesis of CF lung disease into two groups *(12)*. The first describes defects in CFTR that result in altered salt and water concentrations of airway secretions. This then affects host defenses and creates a milieu for infection. The second is associted with CFTR deficiency that results in biologically and intrinsically abnormal respiratory epithelia. These abnormal epithelial cells fail as a mechanical barrier, enhancing the presence of pathogenic bacteria by providing receptors and binding sites or failing to produce functional antimicrobials.

Through multiple mechanisms, the CF patient is compromised in their ability to clear airway colonization and infection. Whereas understanding which immunologic abnormalities are responsible for the disease manifestations leads to more effective therapies, correction of the CFTR defect by gene transfer could ultimately prevent the occurrence or progression of the infectious and inflammatory consequences.

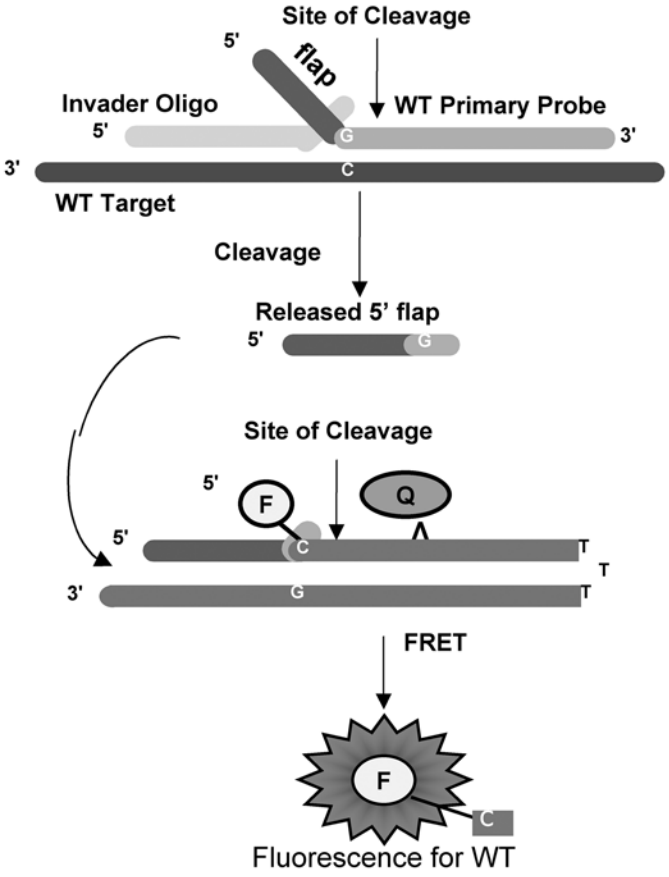


Fig. 1. Schematic illustration of the Invader assay (Third Wave Technologies, Madison, WI).

5. ANATOMIC PATHOLOGY

The pathophysiology of CF leads to clinical, gross, and histologic changes in organs and organ systems expressing abnormal CFTR, including respiratory, pancreas, hepatobiliary, gastrointestinal, and reproductive *(14)*. With the current age of individuals affected with CF ranging from 0 to 74 yr, and the predicted survival age for a newly diagnosed child as 33.4 yr, pulmonary and extrapulmonary disorders (gastrointestinal, hepatobiliary, vascular, and musculoskeletal) will become increasingly manifest *(15)*. The degree of disease correlates with the degree of CFTR function.

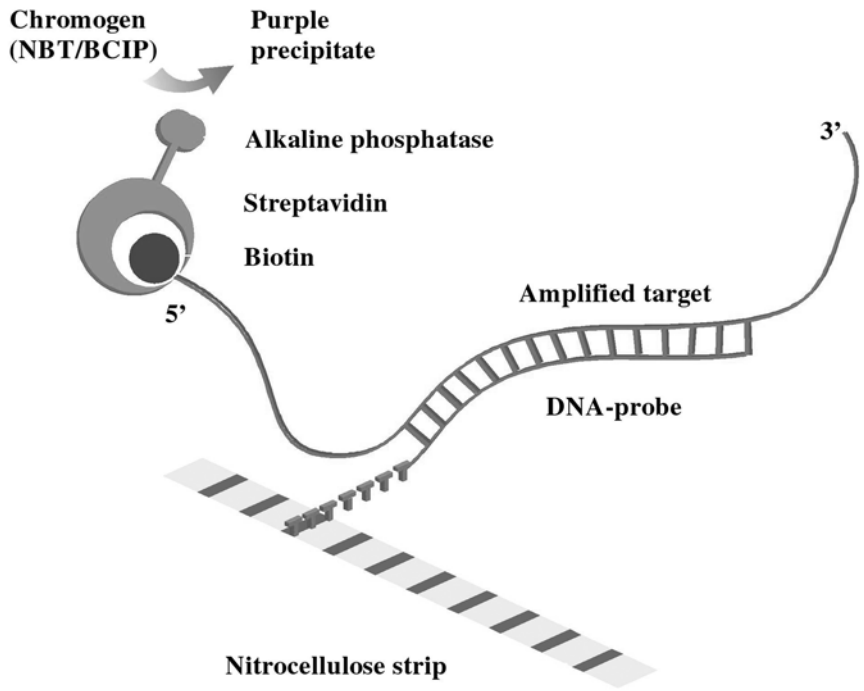


Fig. 2. Schematic representation of a reverse line probe assay.

5.1. RESPIRATORY SYSTEM Lung infection remains the leading cause of morbidity and mortality in CF patients *(16)*. It is currently recognized that CF lung disease is the consequence of recurrent chronic pulmonary infection by the well-known opportunistic pathogens *Pseudomonas aeruginosa* (mucoid and nonmucoid), *Burkholderia cepacia*, *Staphylococcus aureus*, and *Haemophilus influenza (17)*. Recurrent infection and the inflammatory response result in progressive irreversible lung damage, of which bronchiectasis is the landmark change. Bronchial mucous plugging as a result of tenacious mucus facilitates colonization by micro-organisms *(18)*. Often, subpleural bronchiectatic cavities develop and communicate with the subpleural space with resultant spontaneous secondary pneumothorax, the incidence of which increases later in life *(15)*.

5.2. PANCREAS Exocrine pancreas insufficiency is present in the majority of patients with CF. This clinically manifests by failure to thrive and fatty bulky stools because of deficiency of pancreatic enzymes. The pancreatic lesions, however, vary greatly in severity and might be absent in some patients who die in infancy *(19)*. Early in the postnatal development of the pancreas, patients with CF have a deficiency of normal acinar development. Increased secretion of tenaceous and viscous mucus within the ducts and increased duct volume contribute to progressive degradation and atrophy of pancreatic acini. These factors result in duct obstruction and progressive pancreatic pathologic changes *(20,21)*. Exocrine pancreatic disease appears to develop as a result of ductal mucus accumulation resulting from decreased anion secretion. Coupled to normal protein load derived from acinar cell secretion, this leads to pancreatic protein hyperconcentration within the pancreatic ducts. The protein hyperconcentration increases susceptibility to precipitation and duct luminal obstruction *(22–24)*. Hence, the characteristic lesion is cystic ductal dilation, atrophy of pancreatic acini, and severe parenchymal fibrosis.

5.3. HEPATOBILIARY SYSTEM The manifestation of CF in the hepatobiliary system is directly related to CFTR expression. This results in the only inherited liver disease resulting form impaired secretory function of the biliary epithelium *(25)*. Males are more likely to be affected than females and the risk for developing liver disease is beween 4% and 17% as assessed by yearly exams and biochemical testing *(26,27)*. Although a variety of liver manifestations exist *(28,29)*, including fatty infiltration (steatosis), common bile duct stenosis, sclerosing cholangitis *(25,30)*, and gallbladder disease *(31–33)*, the rare but characteristic liver lesion in CF is focal biliary cirrhosis *(34)*. The latter develops in a minority of patients, usually older children and adults, and results from abnormal bile composition, reduced flow, and intrahepatic bile duct obstruction *(35)*.

5.4. GASTROINTESTINAL SYSTEM The gastrointestinal manifestations of cystic fibrosis, seen mainly in the neonatal period, include meconium ileus, distal intestinal obstruction syndrome (DIOS), fibrosing colonopathy, strictures, gastroesophageal reflux, rectal prolapse, and constipation in later childhood *(26,36–40)*. Throughout the intestines, CFTR is the determinant of chloride concentration and secondary water loss into the intestinal lumen. Decreased water content results in viscous intestinal contents, with a 10–15% risk of meconium ileus in babies born with cystic fibrosis and accounting for DIOS and constipation in older children *(41)*. DIOS (formerly meconium ileus equivalent) is a recurrent partial or complete obstruction of the intestine in patients with CF and pancreatic insufficiency *(40)*.

5.5. BONES AND JOINTS Arthritis is a rare but recognized complication of cystic fibrosis that generally occurs in

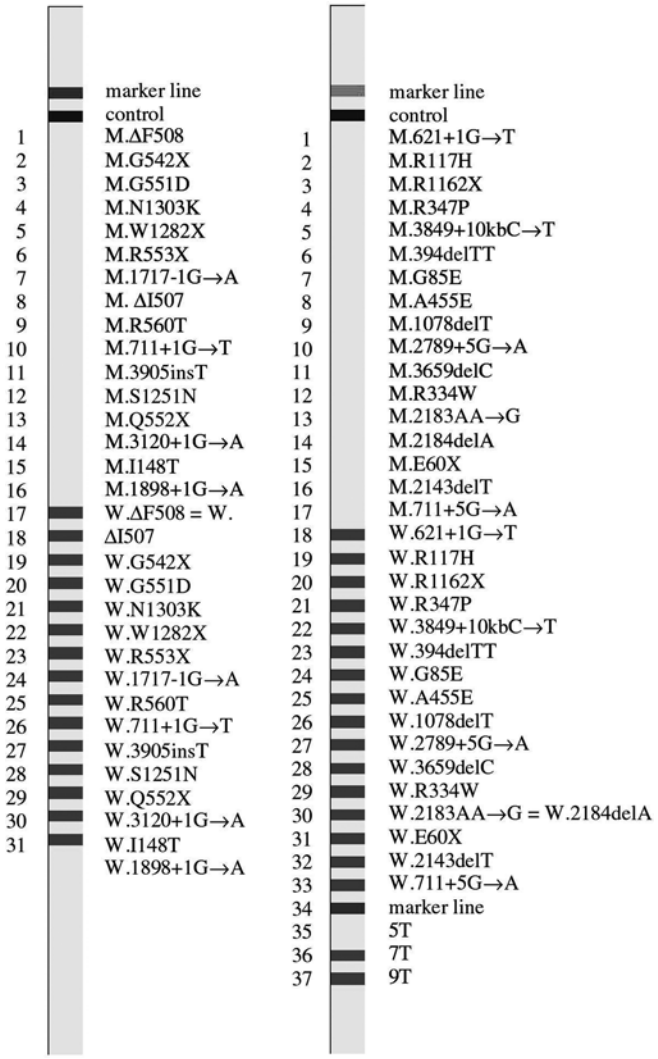


Fig. 3. Innogenetics line probe assay (LiPA, two strips) for 33 CFTR mutations. Top portion of strip contains mutant specific probes and bottom of strip contains normal probes.

the second decade *(42–45)*. Three types of joint disease are described in patients with cystic fibrosis: (1) cystic fibrosis arthritis (CFA) or episodic arthritis (EA), (2) hypertrophic pulmonary osteoarthropathy (HPOA), and (3) coexistent or treatment-related arthritis *(42,43,46,47)*. The most common form, EA, is characterized by an episodic, self-limited polyarticular arthritis with no evidence of progression to joint damage *(42)*. Histologic features are minimal but might include prominent blood vessels and interstitial edema or, rarely, lymphocytic inflammation *(48)*. Immunofluoresence studies have shown deposition of IgM, IgG, and/or C3d in the blood vessels *(43)*.

5.6. REPRODUCTIVE SYSTEM Infertility, an inevitable consequence of cystic fibrosis in males and occurring in greater than 95% of patients, is the result of congenital bilateral absence or atrophy of the vasa deferentia (CBAVD) and/or dilated or absent seminal vesicles *(49)*. Mutations in the CFTR gene are present in 70% of the patients with CBAVD *(50)*. Obstructive azoospermia could be diagnosed by semen analysis; however, it must be confirmed by testicular biopsy and exclusion of other reasons for azoospermia *(51)*. Females might have impaired fertility resulting from dehydrated cervical mucus, yet their reproductive function is normal *(52)*. Advances in techniques such as microscopic epididymal sperm aspiration (MESA) and intracytoplasmic sperm injection have allowed males with cystic fibrosis to reproduce *(16)*.

5.7. VASCULAR SYSTEM Vasculitis of the leukocytoclastic type represents a rare complication for CF patients. Histologic features include transmural infiltration of neutrophils, endothelial swelling, and fibrinoid necrosis. Antineutrophil cytoplasmic autoantibody (ANCA) is present in the sera of patients with CF *(53)*, and autoantibodies to bactericidal/permeability-increasing protein (BPI), a recently characterized ANCA specificity, is present in the sera of CF patients (IgA-anti BPI [83%], and IgG-anti-BPI [91%]) *(54)*.

6. MOLECULAR DIAGNOSTICS

Cystic fibrosis mutation analysis is useful in a variety of clinical settings including (1) diagnosis, (2) newborn screening, (3) prenatal diagnosis for at-risk pregnancies, and (4) carrier detection. The majority of CF mutation analyses are for carrier detection and risk revision. Mutation analysis is a useful adjunct to sweat testing for diagnosis of CF, particularly for patients with borderline sweat tests, patients with atypical clinical presentations with normal sweat electrolytes, and at-risk newborns for whom sufficient quantities of sweat cannot be collected. Mutation analysis is also useful for a diagnosed proband in order to identify the familial mutation for carrier risk revision for at-risk relatives. With respect to newborn screening, both genetic and sweat testing are used after measuring immunoreactive trypsinogen.

Population screening for the purpose of identifying CF mutation carriers has begun in the United States. Using available data from over 20,000 CF patients, the ACMG (American College of Molecular Genetics) and ACOG (American College of Obstetrics and Gynecology) recommended a pan-ethnic panel of 25 mutations that occur at a frequency >0.1% in any of the major US ethnic groups and 4 reflex sequence variants *(55)* (Table 1).

As an established recommendation, CF carrier screening requires newer methods and technologies to support such a national program. Several assays are commercially available as analyte-specific reagents (ASRs) that utilize different technologies (Table 2). All of these assays include at least the ACMG/ACOG minimum core mutation panel. However, each assay/technology varies considerably with respect to criteria that laboratories consider in choosing a diagnostic platform, such as reagent/royalty costs, footprint, throughput, flexibility, and data analysis. PCR royalty payments are required of all assays except the Invader® platform (Third Wave Technologies, Madison, WI). As an enzymatic reaction (Fig. 1), the Invader assay requires only pipetting steps and is easily automated using a liquid handler. This assay should be more cost-effective than other ASRs because it eliminates an amplification step and uses a detection system that requires only a plate reader. For all other ASRs, reagent costs per patient are significant as are labor costs. This latter issue is being addressed by semiautomated platforms that allow for greater throughput and less "hands-on" time.

Two of the more commonly used assays are based on reverse blotting technologies (Fig. 2). The Linear Array CF Gold 1.0

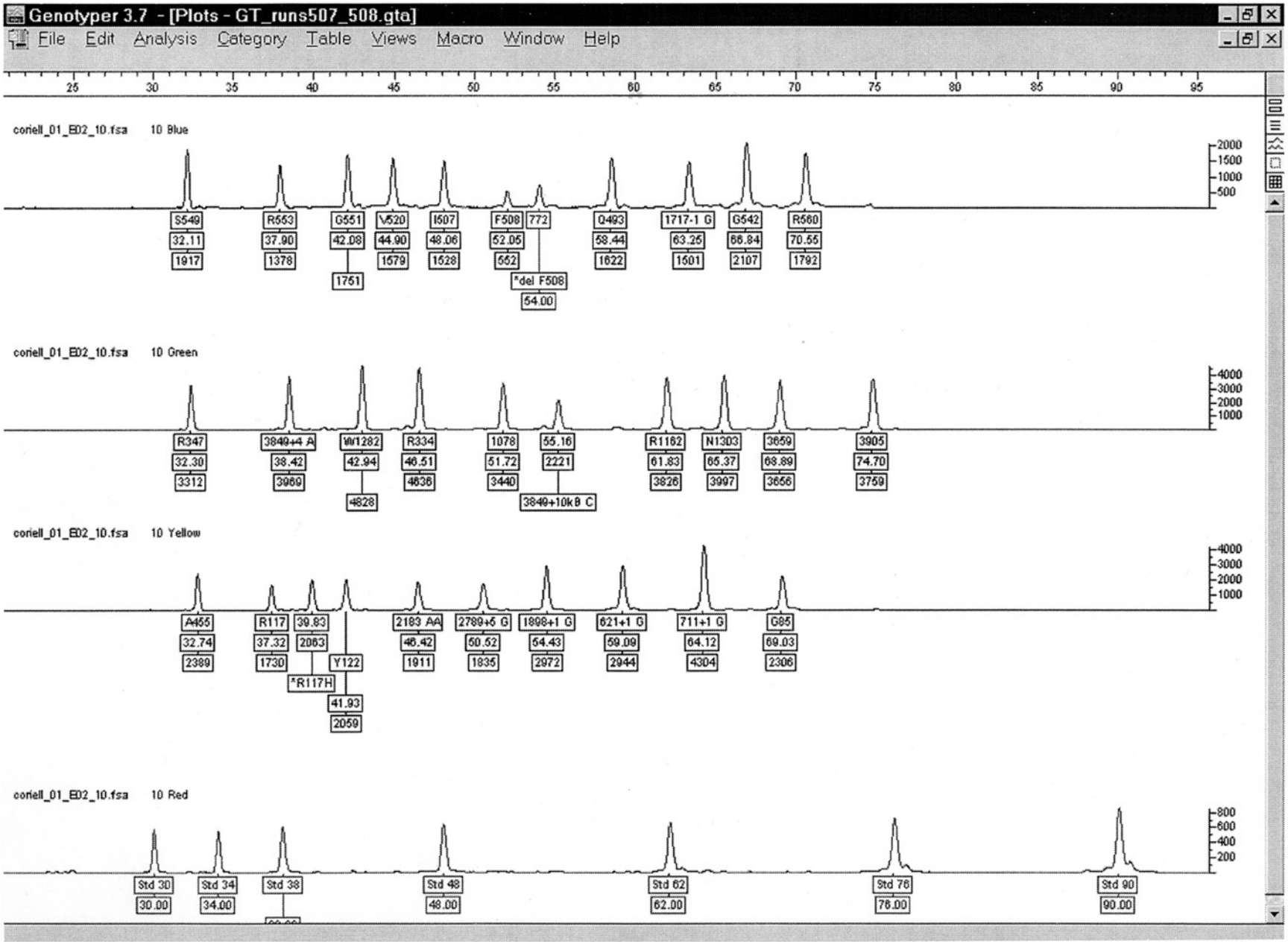


Fig. 4. Representation of the ABI OLA Assay (ABI, Foster City, CA).

Table 2
PCR-Mediated ASRs *for Cystic Fibrosis Carrier Screening*

ASR	*Detection*	*Mutations*	*Automation*	*Company*
CF Gold 1.0	Reverse line blot; colorimetric	25	No	Roche Molecular Systems
Probe Array CFTR33	Line probe assay (LiPA); colorimetric	33	Yes	Innogenetics
ABI CF 2.0	OLA; fluorescent	31	Yes	Applied Biosystems
NanoChip® CFTR	Microarray; fluorescent	25	Yes	Nanogen

Fig. 5. The NanoChip workstation and CF data output (Nanogen Inc., San Diego, CA).

(Roche Molecular Biochemicals, Indianapolis, IN) and the Probe Array CFTR33 (Innogenetics, Inc., Alpharetta, GA) currently screen for 25 and 33 CF mutations, respectively, while also providing optional intron 8 5/7/9T typing (Fig. 3). These reverse line blots or line probe assays (LiPA) are in widespread use because of early availability, ease of use, and ease of validation in the laboratory. The Innogenetics LiPA is also available in a semiautomated platform that minimizes hands-on time for the hybridization and detection steps. Although both assays are amenable to low and high testing volumes, the major disadvantage of these systems is that they are closed platforms, so that additional mutations cannot be tested for easily.

A multiplex oligonucleotide ligation assay (OLA) (ABI, Foster City, CA) and microarray assay (NanoChip®, Nanogen, San Diego, CA) have also been introduced as ASRs. The Nanogen and ABI platforms both offer semiautomated detection and data analysis. Each requires a capital expenditure/reagent rental for instrumentation. The assay formats for both are flexible and can be customized to accommodate additional mutations. Both are amenable to high throughput and can be used as a consolidated work platform for many other assays. OLA is a single - tube non-isotopic test that utilizes automated capillary electrophoresis instrumentation for allele discrimination (Fig. 4). The NanoChip technology offers the latest in allele-specific oligonulceotide hybridization and detection utilizing electronic microarray and fluorescent detection methods (Fig. 5).

7. RESULT REPORTING

Mutation analyses have only three possible results: (1) no mutation detected, the patient is normal at a specific allele; (2) one mutation detected, the patient is heterozygous or a carrier of the mutation; or (3) two mutations detected, the patient is homozygous or genetically affected. However, reporting of CF results is more complex because not all CFTR mutations are tested for and different interpretations for each of the possible results exists depending on the indication for the test and the ethnicity and family history of the consultand.

For example, in a screening program, the finding of one mutation in an asymptomatic patient demonstrates that the patient is a carrier but has a very different interpretation in a diagnostic context. For a symptomatic patient, detection of one mutation could mean that the patient is indeed a CF carrier; however, a second possibility is that the patient might also be a compound heterozygote for one identified mutation and a second, unidentified mutation that is either private or rare. In a diagnostic setting, the finding of two mutations confirms a diagnosis of CF. However, genotype–phenotype relationships are not well enough established that prognostic statements can be made for the many different possible homozygous and compound heterozygous genotypes. The interpretation of one or no detected mutations for a symptomatic patient must include a recommendation for diagnostic sweat electrolyte analysis.

Most CF testing referrals are for carrier screening. Although the finding of one mutation in this context is straightforward, the majority of these analysis are negative—no mutations detected. In these cases, the physician must be informed that the revised carrier risk, although reduced from the prior risk, is a nonzero number because detection is incomplete (i.e., not all mutations tested for). Prior and revised carrier risks are both dependent on patient ethnicity and family history. Prior carrier risks for patients with a negative family history are calculated from the ethnic frequency of CF. Prior carrier risks for patients with a positive family history are based on pedigree analysis. Thus, responsible reporting for CF mutation analysis incorporates the indication for the test, patient ethnicity, and family history. Model laboratory reports can be found in the published ACMG/ACOG recommendations for detection of no or one mutation in a carrier context for individuals with a negative personal and family history *(56)*.

8. CONCLUSION

Cystic fibrosis represents an extremely heterogeneous disease that affects most major organ systems. In addition, our knowledge of the disease mechanisms and genetics makes CF one of the most understood complex diseases at all fundamental levels. CF carrier screening is the first clinically recommended DNA screening program in the United States. In addition to diagnostic and technical challenges, our ability to provide quality education for health care providers and patients with respect to the nuances of genetic testing will determine future progress in carrier screening analyses.

REFERENCES

1. Brennan, A. L., and Geddes, D. M. Cystic fibrosis. Curr. Opin. Infect. Dis.**15:**175–182, 2002.
2. Rommens, J. M., Iannuzzi, M. C., Kerem, B., et al. Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* **245(4922):**1059–1065, 1989.
3. Riordan, J. R., Rommens, J. M., Kerem B., et al. Identification of the cystic fibrosis gene: cloning and characterization of complimentary DNA. *Science* **245(4922):**1066–1073, 1989.
4. Kerem, B., and Rommens, J. M., Buchanan, J. A., et al. Identification of the cystic fibrosis gene: genetic analysis. *Science* **245:** 1073–1080, 1989.
5. Rosenstein, B. J., and Zeitlin, P. L. Cystic fibrosis. *Lancet* **351:** 277–282, 1998.
6. The Cystic Fibrosis Genetic Analysis Consortium. Cystic Fibrosis Mutation Data Base. Available at http://www.genet.sickkids.on.ca/cftr/.
7. Vankeerberghen, A., Cuppens, H., and Cassiman, J. J. The cystic fibrosis transmembrane conductance regulator: an intriguing protein with pleiotropic functions. J. *Cystic Fibrosis* **1**:13–29, 2002.
8. Zielenski, J. Genotype and phenotype in cystic fibrosis. *Respiration* **67**:117–133, 2000.
9. The Cystic Fibrosis Genotype–phenotype Consortium.Correlation between genotype and phenotype in cystic fibrosis. *N. Engl. J. Med.* **329:**1308–1313, 1993.
10. Morral, N., Betranpetit, J., Estivill, X., et al. The origin of the major cystic fibrosis mutation (delta F508) in European populations. *Nat. Genet* **7:**169–175, 1994.
11. Wine, J. J. The genesis of cystic fibrosis lung disease. *J. Clin. Invest*.**103(3):**309–312, 1999.
12. Bales, R., Weiner, D. J., and Wilson J. M. The innate immune system in cystic fibrosis lung disease. *J. Clin. Invest*. **103(3):**303–307, 1999.
13. Goldman, M. J., Anderson, G.M., Stolzenberg, E. D., et al. Human beta-defensin-1 is a salt –sensitive antibiotic in lung that is inactivated in cyctic fibrosis. *Cell* **88:**553–560,1997.
14. Lewis, M. J., Lewis, E. H., III, Amos, J. A., and Tsongalis, G. J. Cystic fibrosis. *Am . J. Clin. Pathol. Pathol.* Patterns, **120** (suppl): s3–13, 2003.
15. Cystic Fibrosis Foundation, Patient Registry. *2001 Annual Report*. Cystic Fibrosis Foundation, Bethesda, MD, 2002.
16. Ratjen, F., and Doring, G. Cystic fibrosis. *Lancet* **361:** 681–689, 2003.
17. Sly, P. and Stick, S. Cystic fibrosis. *Thorax* **56(3):**237–241, 2001.
18. Reyes-Mugica, M. and Gonzalez-Crossi, F. Pulmonary disease in the pediatric age group, in *Pathology of Pulmonary Disease*, Saldana, M. J., ed., Lippincott, Philadelphia, pp. 93–99, 1994.
19. King, A., Mueller, R. F., Heeley, A. F., et al. Diagnosis of cystic fibrosis in premature infants. *Pediatr. Res.* **20(6):** 536–541, 1986.
20. Liu, P., Daneman, A., Stringer, D. A., and Durie, P. R. Pancreatic cysts and calcification in cystic fibrosis. *J. Can. Assoc. Radiol*. **37(4):**279–282, 1986.
21. Imrie, J., Fagan, D., and Sturgess, J. Quantitative evaluation of the development of the exocrine pancreas in CF and control infants. *Am. J. Pathol.* **95(3):**697–707, 1979.
22. Kopelman, H. R., Durie, P., Gaskin, K. J., Weizman, Z., and Forstner, G. G. Pancreatic fluid and protein hyperconcentration in cystic fibrosis. *N. Engl. J. Med.* **312(6):** 329–334,1985.
23. Kopelman, H. R., Forstner, G. G., Durie, P., and Corey, M. Origins of chloride and bicarbonate secretory defect in the cystic fibrosis pancreas as suggested by pancreatic function studies in control and CF subjects with preserved pancreatic function. *Clin. Invest. Med.* **12(3):** 207–211, 1989.
24. Durie, P. R., and Forstner, G. G. The exocrine pancreas, in *Cystic Fibrosis in Adults*. Lippencott-Raven, Philadelphia, PA, pp. 261–287, 1999.
25. O'Brien, S., Keogan, M., Casey, M., et al. Biliary complications of cystic fibrosis. *Gut* **33:**387–391, 1992.
26. Modolell, I., Guarner, L., and Malagelada, J. R. Digestive system involvement in cystic fibrosis. *Pancreatlolgy* **2:**12–16, 2002.
27. Scott-Jupp, R., Lama, M., and Tanner, M. S. Prevalence of liver disease in cystic fibrosis. *Arch. Dis. Child.* **66:**698–701, 1991.
28. Colombo, C., Battezzati, P. M., and Podda, M. Hepatobiliary disease in cystic fibrosis. *Semin. Liver Dis*. **14:**259–269, 1994.
29. Cannon, J. J. Gallbladder and biliary tract disease in cystic fibrosis. *Semin. Liver Dis.* **14:**259–269, 1994.
30. Strandvik, B., Hjelte, L., Gabrielson, N., et al. Sclerosing cholangitis in cystic fibrosis. *Scand. J. Gastroenterol.* **23(Suppl.143):**121–124, 1998.
31. Isenberg, J. N. Cystic fibrosis: its influence on the liver, biliary tree, and bile salt metabolism. *Semin. Liver Dis.* **2:**302–313, 1982.
32. Lambert, J. R., Cole, M., Crozier, D. N., and Cannon, J. J. Intra pancreatic common bile duct compression causing jaundice in an adult with cystic fibrosis. *Gastroenterology* **80:**169–172, 1981.
33. Strandvik, B., Hjelte, L., Gabrielson, N., and Glaumann, H. Sclerosing cholangitis in cystic fibrosis. *Scand. J. Gastroenterol.* **143(Suppl):**121–124, 1988.
34. Stern, R. C., Rothstein, F. C., and Doershuk, C. F. Treatment and prognosis of symptomatic gallbladder disease in patients with cystic fibrosis. *J. Pediatr. Gastroenterol. Nut* **5:**35–40, 1986.
35. Colombo, C., Apostolo, M. J., Ferrari, M., et al. Analysis of risk factors for the development of liver disease associated with cystic fibrosis. *J. Pediatr* **124:**129–132, 1994.
36. Hausler, M., Heimann, G., and Biesterfeld, S. Fibrosis: a general feature in cystic fibrosis? *J. Pediatr. Gastroenterol. Nut* **34:**236–239, 2002.
37. Rubenstein, S., Moss, R., and Lewiston, N. Constipation and meconium ileus equivalent in patients with cystic fibrosis. *Pediatrics* **78:**473–479, 1986.
38. Pawel, B. R., Chadarevian, J. P., and Franco, M. E. The pathology of fibrosing colonopathy of cystic fibrosis: a study of 12 cases and review of the literature. *Hum. Pathol.* **28:**395–399, 1997.
39. Smyth, R. L., Ashby, D., O'Hea, U., et al. Fibrosing colonopathy in cystic fibrosis: results of a case-control study. *Gastroenterlogy* **111:**260–264, 1996.
40. Lewis, T. C., Casey, S. C., and Kapur, R. P. Clinical pathologic correlation: a 3-year old boy with cystic fibrosis and intestinal obstruction. J. *Pediatr.* **134:**514–519, 1999.
41. Milla, P. J. Cystic fibrosis: present and future. *Digestion* **59:** 579–588, 1998.
42. Bourke, S., Rooney, M., Fitzgerald, M., et al. Episodic arthropathy in adult cystic fibrosis. *Q. J. Med.* **244:**651–659, 1987.
43. Johnson, S., and Knox, A. J. Arthropathy in cystic fibrosis. *Respir. Med.* **88:**567–570, 1994.
44. Newman, A. J., and Ansell, B. A. Episodic arthritis in children with cystic fibrosis. *J. Pediatr.* **94:**594–596, 1979.
45. Schidlow, D.V., Goldsmith, D. P., Palmer, J., et al. Arthritis in cystic fibrosis. *Arch. Dis. Child.* **59:**377–379, 1984.
46. Turner, M. A., Baildam, E., Patel, L., et al. Joint disorders in cystic fibrosis. *J. R. Soc. Med.* **90(Suppl. 31):**13–20, 1997.
47. Pertuiset, E., Menkes, C. J., Lenoir G., et al. Cystic fibrosis arthritis. A report of five cases. *Br. J. Rheumatol.* **31:**535–538, 1992.
48. Phillips, B.M., and David, T. J. Pathogenesis and management of arthropathy in cystic fibrosis. *J. R. Soc. Med.* **79(Suppl.12):**44–50, 1986.
49. Dodge, J. A. Male fertility in cystic fibrosis. *Lancet* **346:**587–588, 1995.
50. Chillon, M., Casals, T., Mercier, B., et al. Mutations in the cystic fibrosis gene in patients with congenital absence of the vas deferens. *N. Engl. J. Med.* **332:**1475–1480, 1995.
51. Robert, S. C. Current concepts: the diagnosis of cystic fibrosis. *N. Engl. J. Med.* **336:**487–491, 1997.
52. Kopito, L. E., Kosasky, H. J., and Shwahman, H. Water and electrolytes in cervical mucus from patients with cystic fibrosis. *Fertil. Steril.* **24:**512–516, 1973.
53. Ratjen, F., and Doring, G. Cystic fibrosis. *Lancet* **361:**681–689, 2003.
54. Finnegan, M. J., Hinchcliffe, J., Russell-Jones, D., et al. Vasculitis complicating cystic fibrosis. *Q. J. Med.* **267:**609–621, 1989.
55. Richards, C. S., Bradley, L. A., Amos, J., et al. Standards and guidelines for CFTR mutation testing. *Genet. Med* **4:**379–391, 2002.
56. *Preconception and prenatal carrier screening for cystic fibrosis.* American College of Obstetricians and Gynecologists/American College of Medical Genetics. 2001, pp.1–31.

28 Prenatal Genotyping for Identification of Fetuses at Risk for Immune Cytopenic Disorders

MARTIN J. HESSNER AND BRIAN R. CURTIS

1. INTRODUCTION

The genetic characterization of many blood antigen systems has enabled the development of molecular diagnostic assays capable of providing physicians with information that can greatly improve patient care. This technology, when applied to the field of obstetrics, is useful in identifying those fetuses at risk for immune cytopenic disorders. In these disorders, the mother lacks a genetic marker that the child might inherit from the father. Offspring of such parents, therefore, have at least a 50% chance of inheriting the paternal alloallele to which the mother might be immunologically sensitized. If the mother has been previously sensitized to this paternal antigen, either through a previous pregnancy or blood transfusion, the fetus will be at risk for the immune cytopenic disorder, because maternal antibodies can cross the placenta and mediate destruction of the targeted cell. Fetuses that do not inherit the alloallele will not be at risk.

Fetomaternal incompatibilities involving several platelet alloantigens (Table 1) might result in neonatal alloimmune thrombocytopenic purpura (NATP), a condition in which maternal IgG antibodies cross the placenta and destroy fetal platelets, predisposing the fetus to bleeding, possible brain damage, or even death. Incompatibilities within the Fcγ III receptor b (CD16) and HNA-2 (CD177) antigen might result in neonatal alloimmune neutropenia (NAN), a condition in which the fetal neutrophils are destroyed as a result of a maternal IgG antibody, leaving the newborn susceptible to infection. Hemolytic disease of the newborn (HDN), the focus of this chapter, can occur when there exists fetomaternal incompatibilities within any number of different erythrocyte antigen systems, including the RhD, RhCc, RhEe, Kell, Kidd, and Duffy antigen systems. HDN results in the destruction of fetal erythrocytes as a result of the presence of a maternal IgG antibody. Permanent neurological damage can be a result of HDN, and in extreme cases, loss of the fetus or death of the neonate could occur. Investigative and therapeutic measures used for alloimmunized pregnant women involve some risk to the fetus. Currently, women who present with alloantibody titers to red cell antigens will be monitored by amniotic fluid spectrophotometric analysis to detect deviation from linearity at 450 nm, the wavelength at which bilirubin absorbs *(1)*. Accurate determination of fetal risk is achieved through serial analysis, generally weekly for several weeks. Although the risk of placental trauma during amniocentesis has been greatly reduced since the introduction of ultrasound imaging techniques, there still remains a 2% risk of placental trauma *(2)*. Alternatively, percutaneous umbilical blood sampling allows direct measurements of all fetal blood parameters, including blood groups. However, because of the risk of fetomaternal hemorrhage and further sensitization of the mother, its use is limited. Prenatal identification of the relevant genotypes for fetuses potentially at risk for HDN requires fetal DNA isolated from only a single amniocentesis. This can obviate the need for expensive and invasive monitoring throughout the pregnancy when fetuses are shown to be compatible with sensitized mothers and therefore not at risk. Genetic identification of fetuses at risk, on the other hand, allows for appropriate monitoring and early intervention.

This chapter focuses on typing red cell antigen systems involved with the development of HDN and the caveats associated with this area of molecular diagnostics, namely allelic variants and the contamination of fetal samples with maternal (nucleated blood) cells. Many hematologic antigen systems arise through single-nucleotide polymorphisms (SNPs), and in recent years, numerous technologies have been developed that enable rapid and accurate SNP typing that utilize various allelic discrimination biochemistries, reaction formats, and detection methods. These methods include fluorescent bead-based technologies (Luminex, Illumina, Q-dot), automated enzyme-linked immunosorbant (ELISA) assays (Orchid Biocomputer), fluorescent detection of pyrophosphate release (pyrosequencing), fluorescence resonance energy transfer (FRET)-based cleavase assays (Third Wave Technologies) *(3,4)*, and others. These

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

Table 1
Human Platelet Antigens[a]

Antigens	*Other Names*	*Phenotypic Frequency (HPA)*	*Glycoprotein Location/ Amino Acid Change*	*Nucleotide Substitution*
HPA-la (PI^{A1})	$P1^{A}$, Zw	72% a/a	GPIIIa/Leu:Pro_{33}	T:C 196
HPA-lb (PI^{A2})		26% a/b		
		2% b/b		
HPA-2a (Ko^{b})	Ko, Sib	85% a/a	GPIb/Thr:Met_{145}	C:T 524
HPA-2b (Ko^{a})		14% a/b		
		l% b/b		
HPA-3a (Bak^{a})	Bak, Lek	37% a/a	GPIIb/lie:Ser_{843}	T:G 2622
HPA-3b (Bak^{b})		48% a/b		
		15% b/b		
HPA-4a (Pen^{a})	Pen, Yuk	>99.9% a/a	GPIIIa/Arg:Gln_{143}	G:A 526
HPA-4b (Pen^{b})		<0.1% a/b		
		<0.1% b/b		
HPA-5a (Br^{b})	Br, He, Zav	80% a/a	GPIa/Glu:Lys_{505}	G:A 1648
HPA-5b (Br^{a})		19% a/b		
		l% b/b		
HPA-6bw	Ca^{a}, Tu	<1% b/b	GPIIIa/Arg:Gln_{489}	A:G 1564
HPA-7bw	Mo^{b}	<1% b/b	GPIIIa/Pro:Ala_{407}	G;C 1317
HPA-8bw	Sr^{a}	<0.1% b/b	GPIIIa/Arg:Cys_{636}	T:C 2004
HPA-9bw	Max^{a}	<1% b/b	GPIIb/Val:Met_{837}	A:G 2603
HPA-l0bw	La^{a}	1% b/b	GPIIIa/Arg:Gln_{62}	A:G 281
HPA-llbw	Gro^{a}	<0.5% b/b	GPIIIa/Arg:His_{633}	A:G 1996
HPA-12bw	Iy^{a}	1% b/b	$GPIb_{\beta}$%/G]y:Glu_{15}	A:G 141
HPA-13bw	Sit^{a}	<1% b/b	GPIa/Met:Thr_{799}	T:C 2531
HPA-14bw	Oe^{a}	1%	GPIIIa/	AAG 1929-31
HPA-15a(Gov^{b})	Gov	35% a/a	CD109/Tyr:Ser_{703}	A:C 2108
HPA-15b (Gov^{a})		42% a/b		
		23% b/b		
HPA-I6bw	Duv^{a}		GPIIIa	C:T 517
NA	Nak^{a}	99.8% (Caucasian)	CD36 (GPIV)	T:G 1264
		97% (African)		C:T 478
		96% (Asian)		

[a]Phenotypic frequencies for the antigens shown are for the Caucasian population only. Significant differences in gene frequencies may be found in African and Asian populations.

newer approaches are now beginning to be used in clinical reference laboratories; however, many groups continue to use allele-specific polymerase chain reaction (ASPCR), a sensitive technique for the discrimination of alleles arising from base substitutions or deletions that does not necessitate postamplification restriction endonuclease digestion or hybridization. ASPCR is based on the principle that the specificity of PCR is conferred by the 3′ end of the oligonucleotide primer, therefore, amplification efficiency is greatly reduced when a 3′-terminal mismatch exists between the oligonucleotide primer and the DNA target. Our laboratory has developed, validated, and clinically implemented a number of ASPCR assays that have been useful for identifying fetuses at risk for HDN, NAN, and NATP *(5–14)*.

2. THE RH BLOOD GROUP

The Rhesus (Rh) blood group system contains antigens produced from two distinct genes, *RHD* and *RHCE*, that are tandemly localized on chromosome 1 and are thought to have arisen through duplication of a single ancestral gene *(15,16)*. The *RH* genes are greater than 95% homologous at the nucleotide sequence level, both consist of 10 exons spanning over 75 kb, and both encode peptides of 417-amino-acid residues, with a predicted molecular mass of 30–35 kDa *(17–19)*. Approximately 15% of Caucasians are RhD-negative, where the D-negative phenotype is the result, in general, of absence of the *RHD* gene; such individuals, however, possess two copies of the *RHCE* gene. There are a number of sequence differences between the *RHCE* and *RHD* genes, which can be utilized in genotyping assays to identify the presence or absence of the *RHD* gene as well as detect the RhC/c– and RhE/e–specific polymorphisms within the *RHCE* gene. The Rh antigens are among the most antigenic of the polymorphic red cell surface proteins making them highly clinically significant, especially in the context of HDN *(20)*. HDN resulting from RhD has become relatively infrequent because of the use of Rh immune globulin (RhoGAM), given to RhD-negative mothers giving birth to D-positive infants, thereby passively eliminating circulating and potentially immunogenic fetal red cells before they are recognized by the maternal immune system *(21)*.

2.1. GENOTYPING THE RHD RED CELL ANTIGEN SYSTEM Polymerase chain reaction (PCR)-based RhD typing assays, utilizing sequence differences in intron 4, exon 3, exon 7, and the untranslated region within exon 10, have been previously described *(18,22–25)*. However, discrepancies between serotyping and genotyping have been observed. These

discrepancies are largely the result of the existence of allelic variants, the molecular basis of which is often the result of recombination between the *RHD* and *RHCE* genes. In these hybrid genes, some *RHD* sequences are replaced with the corresponding sequences from the *RHCE* gene *(22–26)*. For other variants, all that is known is that the sequences targeted by PCR primers or restriction enzymes are altered or deleted. RhD genotyping is commonly accomplished by multiplexing the oligonucleotide primers of Arce et al. with those described by Bennett et al *(18,23)*. The RhD genotyping strategy described by Arce et al., detects a 600-bp deletion within intron 4 of the *RHD* gene, which is not present within intron 4 of the *RHCE* gene *(18)*. Amplification with a primer pair that targets exons 4 and 5 of the RhD and the RhCcEe genes results in a 1200-bp RhCcEe product and a 600-bp product in RhD-positive individuals. The RhD-specific oligonucleotide primers described by Bennett et al. specifically amplify a 193-bp product of the 3′ untranslated region of exon 10 *(18,23)*. RhD genotyping can also be performed through specific amplification of a 96-bp product from exon 7 of the *RHD* gene, previously described by Simsek et al. *(25)*. Finally, specific amplification of a 111-bp product from exon 3 of the *RHD* gene can be used for genotyping, as previously described by Beckers et al. *(22)*.

Among Caucasians, a high concordance between serology and genotyping has been observed when methods targeting sequence differences in intron 4 and the untranslated region within exon 10 are combined (Fig. 1); however, false-positive results, especially among African and Asian individuals, have been observed *(22–25)*. Less than 5% of Native South Africans are serologically RhD-negative; however, less than 20% of serologically D-negative individuals have been observed to completely lack the *RHD* gene *(27)*. In 66% of serologically D-negative native Africans, a silent *RHD* gene (*RHDψ*) has been identified that possesses a 37-bp insertion in exon 4, which presumably produces a premature stop codon at position 210, resulting in a truncated protein *(27)*. Discordance between RhD serology and *RHD* genotyping results has been a potential problem for laboratories conducting prenatal genotyping *(28–30)*. Therefore, we evaluated and augmented our standard *RHD* genotyping assay, which specifically targets exon 10 of *RHD* and the intron 4 600-bp deletion/insertion polymorphism of the *RHD/RHCE* genes, with an additional reaction possessing primers that identify the 37-bp exon 4 insertion that gives rise to the *RHDψ* allele *(27)*. Primers flanking the exon 4 insertion point were used for detection of *RHD* and *RHDψ* among a total of 231 serotyped individuals; 134 African-American, 85 Caucasian, and 12 RhD serotype-negative/genotype-positive, D-sensitized women (Fig. 1) *(14)*. *RHDψ* was detected in 19% ($n = 4/21$) of RhD sero-negative African Americans and 4.4% ($n = 5/113$) of RhD sero-positive African-Americans *(14)*. Complete concordance was observed, with this additional primer set, between serology and genotyping when detecting the 381-bp normal *RHD* PCR product, whereas detection of the 418-bp *RHDψ* gene product was useful in resolving 10/12 previously ambiguous prenatal genotypings where the RhD-sensitized mother possessed an apparently intact *RHD* gene in the standard assay *(14)*. In these cases, it was possible, with the *RHDψ* primer set, to determine the basis for the discrepancy between the maternal serotype and genotype and, second, to determine if the fetus had paternally inherited *RHD* or maternally inherited *RHDψ*, or both. The addition of this primer set to *RHD* genotyping strategies enables definitive genotyping of most RhD-negative African-Americans based on the reported frequencies of 54% and 24% for the homozygous deletion of *RHD* and *RHDψ* alleles, respectively, as the genetic basis for D-negativity *(27)*. In our study, 19% ($n = 4/21$) of the RhD sero-negative African-American donors possessed *RHDψ*, which is consistent with previous reports *(14,27)*.

Fifty-six percent of the RhD-positive individuals within the Caucasian population are heterozygous for RhD *(31)*, making knowledge of the father's zygosity useful for predicting the probability of a couple's conception of an RhD-positive pregnancy, especially if the mother has been immunized through a previous pregnancy. Recently, reliable approaches utilizing real-time quantitative PCR and ASPCR have been reported that allow determination of *RHD* gene dosage *(32)*. These assays can reduce the need for amniocentesis, because a homozygous *RHD*-positive male can only produce RhD-positive offspring, making the invasive procedure unnecessary. However, heterozygous males have a 50% chance of fathering a child compatible with an RhD-negative mother.

The desire for noninvasive fetal diagnosis has been encouraged by the detection of fetal cells in maternal circulation *(33–35)* as well as the detection of fetal DNA in maternal plasma *(36,37)*. However, the low numbers of fetal cells in maternal circulation *(38)* and their persistence for up to decades after delivery *(39)* have been an obstacle to their molecular diagnostic utilization. Fetal DNA, on the other hand, has been reported at mean fraction concentrations of 3.4% and 6.2% in maternal plasma during early and late pregnancies, respectively *(40)*, and, furthermore, fetal DNA has also been reported to rapidly clear from maternal plasma, possessing a half-life of only minutes *(41)*. These characteristics make fetal DNA present in maternal plasma an attractive target for prenatal genotyping, and several groups have used this approach for *RHD* genotyping of fetuses carried by alloimunized RhD-negative mothers *(42–45)*. Collectively, their results indicate that reliable typing can be obtained, especially during and after the second trimester. Realizing the full diagnostic potential of this source of fetal DNA will require sensitive and highly discriminating detection chemistries, because many antigen systems involved in immune cytopenic disorders differ by a single nucleotide, unlike RhD; however, the future appears promising.

2.2. GENOTYPING RHCC AND RHEE RED CELL ANTIGEN SYSTEMS The *RHCE* gene encodes the peptide carrying the Rh E/e and C/c antigens *(46)*. The RhC/c epitope arises from a single-base substitution within exon 2 at nt307; this substitution results in the incorporation of serine (RhC = CCT) or proline (Rhc = TCT) at amino acid 103 *(19,47)*. There are five additional base substitutions between the *RHC* and *RHc* alleles within exons 1 and 2; however, these polymorphisms have been shown not to be involved in the RhC/c serology *(48,49)*. The expression of the C/c antigens is thought to be conformation dependent because the RhD peptide also specifies serine at residue 103 but does not express RhC antigenicity. The RhE/e epitope arises from a single-nucleotide substitution within exon 5 at position 676; this substitution results in the

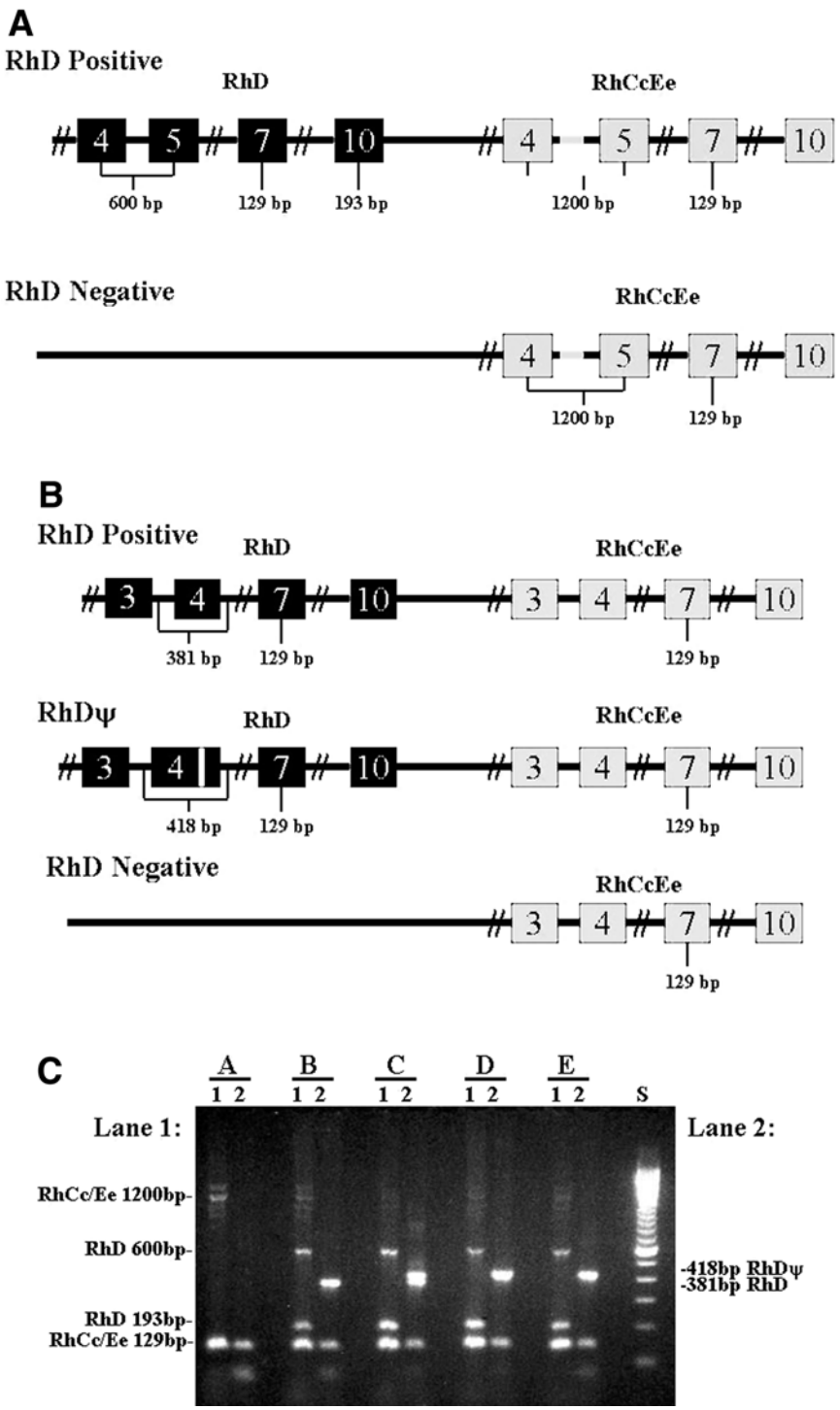


Fig. 1. *RHD* genotyping: **(A,B)** Illustrations of targeted exons and introns of *RHCE, RHD,* and *RHDψ* for *RHD* and *RHDψ* genotyping, respectively *(14,18,23,27)*. **(C)** *RHD* genotyping by ASPCR and 2% agarose gel electrophoresis. Products of each analysis are run in pairs; the intron 4/exon 10 standard *RHD* reaction is in the first lane of each pair and the *RHDψ* reaction is in the second lane of each pair. The *RHD* amplification generates 600-bp and 193-bp *RHD*-specific products, as well as 1200-bp and 129-bp internal control products derived from *RHCE.* The *RHDψ* reaction generates a 381-bp product from a normal *RHD* allele and a 418-bp product from *RHDψ* alleles, a primer pair targeting RhCE provides an internal control *(14,18,23,27)*. Set A: normal RhD-negative individual; set B: normal *RHD*-positive individual; set C: normal *RHD*-positive/*RHDψ* heterozygous individual; panel D/panel E: prenatal genotyping case involving a serologically RhD-negative/standard genotype-positive mother possessing the *RHDψ* allele (panel D), and the initially ambiguous fetal genotyping (panel E). Lane S: 100-bp DNA ladder (Gibco/BRL, Gaithersburg, MD).

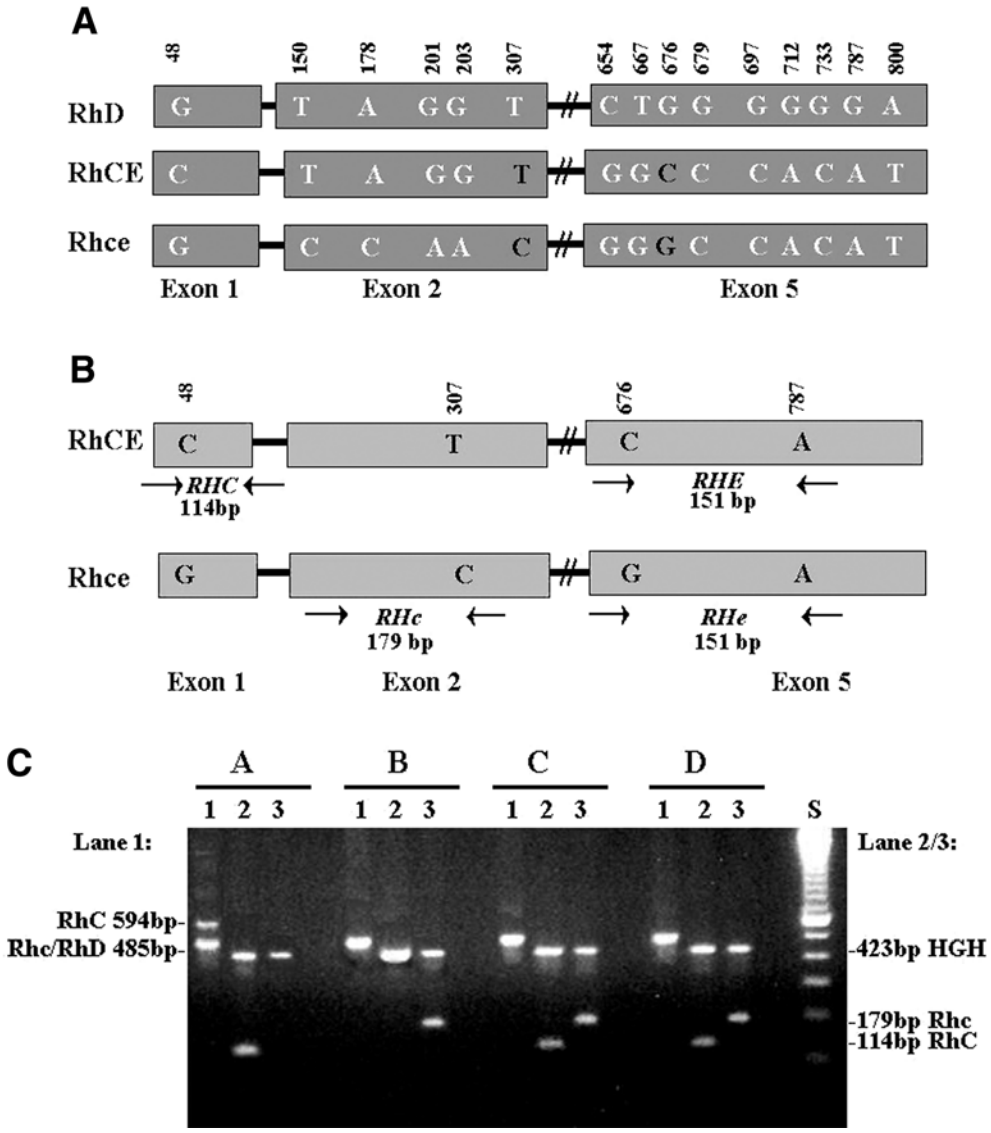


Fig. 2. (A) Nucleotide differences between the *RHCE* and *RHD* genes. Numbers indicate nucleotide positions, and epitope-related positions are illustrated in black. Note that, for simplicity, only *RHCE* and *RHce* haplotypes are shown (*RhcE* and *RHCe* haplotypes exist). **(B)** Strategy for typing *RHCc/Ee* alleles by ASPCR targeting nucleotides 48, 307, and 676. **(C)** *RHC/c* genotyping by ASPCR and 2% agarose gel electrophoresis. Products of each analysis are run in sets of three; lane 1: the *RHC* intron 2 reaction; lane 2: the *RHC* nt48 cytosine reaction; lane 3: the *RHc* nt308 cytosince reaction. The *RhC* intron 2 reaction produces a 594-bp product if the *RhC* allele is present and a 485-bp product from either *RHc* or *RHD*. In the RhC nt48 cytosine reaction and the Rhc nt308 cytosine reaction, a 423-bp HGH control product is detected in each reaction possessing the DNA template, whereas the *RHC* nt48 amplicon generates a 114-bp product and the *RHc* nt 308 amplicon generates a 179-bp product. Panel A: normal RhC homozygotes; panel B: normal Rhc homozygotes; panel C: serological RhCc heterozygote which lacks the intron 2 insertion (RhC false-negative); panel D: serological Rhcc homozygous sample which possesses cytosine at nt48 (RhC false-positive). Lane S: 100-bp DNA ladder (Gibco/BRL, Gaithersburg MD).

incorporation of proline (*RHE* = CCT) for alanine (*Rhe* = GCT) at amino acid 226 *(19,47)*.

The homology between the *RHCE* and *RHD* genes again presents an obstacle when attempting to genotype the RhCc system. However, it is possible to design allele-specific primers for *RHc, RHE*, and *RHe* that target the epitope-specifying nucleotide (Fig. 2). When genotyping *RHc* utilizing nt307 in our laboratory, complete correlation was observed between the presence of cytosine at nt307 and Rhc serology in 282 individuals of Caucasian and African-American ancestry *(14)*. Unfortunately, it is not possible to selectively amplify *RHC* in the presence of *RHD* using the epitope-specific polymorphism at position 307 because the *RHC* allele is identical to RhD within exon 2 *(50,51)*. Hyland et al. *(48)* and Wolter et al. *(49)* reported that the C to G transition at nt 48 within exon 1 correlated well with RhC or Rhc serological typing, respectively; however, approx 5% of the *RHc* alleles mimicked *RHC* in that they possessed cytosine at nucleotide 48. In a study with African blacks, 74% of Rhcc individuals contained a cytosine at nucleotide 48 *(49)*. We have observed a 7.2% (4/55) false-heterozygous rate when genotyping serologically typed Rhcc Caucasian individuals *(14)*, which is consistent with previous reports *(48,49,52,53)*. However, we have also observed a 56.3% (45/80) false-heterozygous rate when genotyping serologically typed Rhcc African-American individuals, which drastically

decreases the utility of this polymorphic position for *RHC* genotyping within this ethnic group. No false-negative genotyping results were observed when using the RhC-associated nt48 cytosine in either African-Americans or Caucasians. Therefore, this locus can be cautiously used to establish the fetal absence of the *RHC* allele in Rhcc mothers sensitized to RhC, with the recognition that there is the possibility of a false-heterozygous genotyping of the fetus.

A unique 109-bp insertion within intron 2 of the *RHCE* gene has been correlated with RhC serology in two independent studies that collectively evaluated over 600 individuals with complete concordance *(52,54)*. In our laboratory, primers flanking the *RHC*-specific intron 2 insertion were used to genotype 282 serotyped individuals (128 African-Americans, 154 Caucasians). Among 154 Caucasian samples, an overall 100% concordance rate was observed between serology and presence or absence of the RhC-associated intron 2 insertion, and complete resolution of the 4 serological Rhcc Caucasian nucleotide 48 false-positive samples was observed *(14)*. Among African-Americans, no false-positives were observed when utilizing the RhC-associated intron 2 insertion, and complete resolution of the 45 serological Rhcc African-American nucleotide 48 false-positive samples was observed. However, a 23.9% (11/46) false-negative rate was detected when genotyping serotyped heterozygotes. It is for this reason this polymorphism cannot be used independently for indirect genotyping of the *RHC* allele; however, this polymorphism can, in certain circumstances, be useful for use in tandem with the exon 1 nt48 cytosine for *RHC* genotyping. Tax et al. has recently reported a new strategy for more accurate genotyping *RHC* alleles through selective amplification of *RHC* variants and has observed >99% concordance with serology across 1071 individuals of different racial/ethnic groups *(55)*. The RhEe system arises through a single-point mutation within exon 5 (Fig. 2) and genotyping can be accomplished without complications arising from RhD *(50,51)*.

The Rh system is complex and highlights some of the caveats associated with using DNA-based reference tests. In order to identify cases where allelic variants exist, regardless of the antigen system in question, it is advisable to always establish concordance between the serotypes and genotypes of the parents before predicting fetal phenotype based on a fetal genotyping result.

3. GENOTYPING THE KELL RED CELL ANTIGEN SYSTEM

The Kell system, a major antigenic system in human erythrocytes, possesses more than 20 different related antigens *(56,57)*. The Kell antigens are localized on a 93-kDa red blood cell integral membrane glycoprotein *(56,57)*. K1 and K2 are alternate, autosomally inherited, codominant alleles. The *KEL* gene is located on chromosome 7 (q33), spans approx 21.5 kb, contains 19 exons, and predicts a peptide of 731-amino-acid residues *(58,59)*. At the molecular level, the difference between the K2 and K1 alleles is a single-base change within exon 6 at position 698, which results in the incorporation of methionine (ATG) for threonine (ACG) at amino acid residue 193 *(58,59)*. K1 is an effective immunogen, being second after RhD in immunogenicity relative to other blood group antigens, and K1 antibodies can cause severe reactions to transfusion of incompatible blood, as well as hemolytic disease of the newborn. The Kell phenotype frequencies in Caucasians are 0.2%, 8.8%, and 91.0% for K1/K1, K1/K2, and K2/K2, respectively, and K1 is observed in approx 4% of the black population *(60)*. Maternal sensitization to K1 is observed in 0.1% of all pregnancies, making anti-K1 the fourth most common antibody encountered in prenatal testing *(61)*.

Although the administration of Rh immune globulin has led to a decrease in the number of mothers sensitized to the RhD antigen, there is no analogous treatment to prevent maternal sensitization to K1. However, hemolytic disease arising from maternal antibodies to other fetal antigen systems, including those of the Kell system, can be as severe as hemolytic disease resulting from anti-D antibodies. Like the RhEe system, genotyping for the *KEL1* and *KEL2* alleles by ASPCR is relatively straightforward, and reliable assays have been published *(12,62)*.

4. GENOTYPING THE KIDD RED CELL ANTIGEN SYSTEM

The Kidd blood group antigen system on human erythrocytes is defined by two alternate codominant antithetical specificities, Jk^a and Jk^b *(60)*. The Kidd antigens are localized on a 43-kDa red blood cell integral membrane protein that functions as a urea transporter *(63–67)*. The *JKA* and *JKB* allele frequencies in Caucasians are 0.51 and 0.49, respectively, giving rise to the three common phenotypes Jk(a+b–), Jk(a–b+), and Jk(a+b+). The Jk(a–b–) phenotype, which is very rare in Caucasians, is observed in 0.1–1.4% of Polynesian populations *(60)*. The *KIDD*/urea transporter gene (*HUT11*) is located on chromosome 18 (q12-q21) and encodes a peptide of 391-amino-acid residues *(64–67)*. The Jk^a and Jk^b epitopes arise through a single G→A transition at nucleotide position 838, which results in the incorporation of aspartic acid or asparagine at amino acid residue 280, respectively *(67)*. ASPCR has been used by our laboratory to define *JKA* and *JKB* by detecting the G→A transition within codon 280 (Fig. 3) *(10)*. We have also observed that Polynesian individuals possessing the recessive silent Jk(a–b–) phenotype genotype as *JKB* homozygotes, indicating that the null allele responsible for this phenotype was derived from *JKB*, and this allele would generate false *JKB*-positive results in our genotyping assay. The molecular mechanisms responsible for the Jk(a–b–) phenotype have been described *(63,68)*. In Polynesian Jk(a–b–) individuals, a G→A 3′ splice acceptor site mutation exists within intron 5, which results in skipping of exon 6, whereas in Finnish Jk(a–b–) individuals, another point mutation within exon 9 (T871C) predicts the loss of a potential glycosylation site *(63)*. To determine the potential impact of the more common Polynesian null *JK* allele on prenatal genotyping, 753 unrelated subjects of different racial/ethnic groups were screened using ASPCR to detect the intron 5 G→A 3′ splice acceptor site mutation *(69)*. The intron 5 G→A 3′ splice acceptor mutation responsible for the Polynesian Kidd (a–b–) phenotype was not observed in African-Americans, Caucasians, Hispanics, Native Americans, or Koreans screened, whereas one heterozygous carrier was observed among 91 Asian Indians *(69)*. Genotyping the heterozygous individual for the Jk^a/Jk^b dimorphism at nt838 revealed only the presence of "G," indicating a single normal *JKA* allele and a single null *JKA* allele. Unfortunately, fresh blood for serological analysis of this Asian Indian individual was

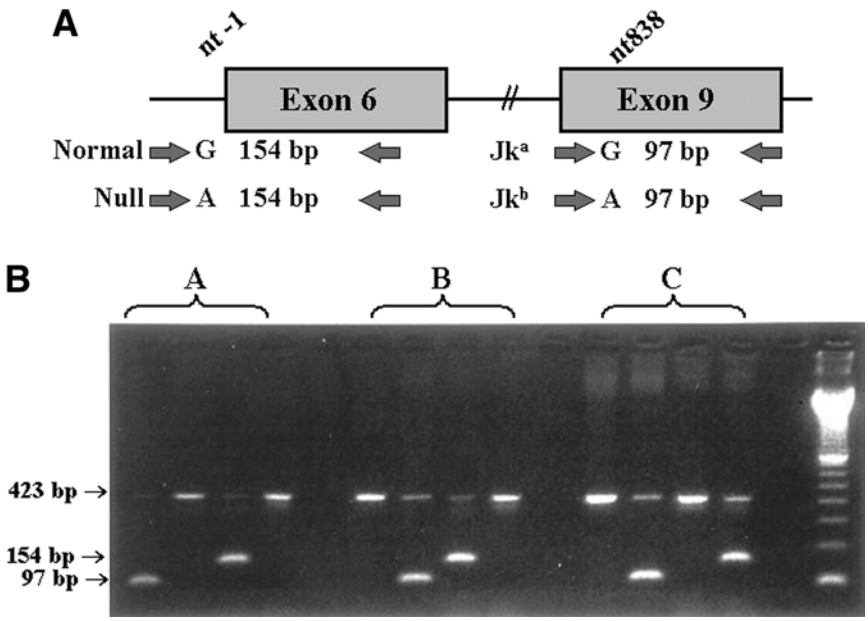


Fig. 3. Kidd genotyping by ASPCR. (**A**) Schematic of polymorphisms associated with normal/null and *JKA* and *JKB* Kidd alleles. (**B**) Products of each analysis are run sets of four ordered as follows: the *JKA* (nt838G) reaction, the *JKB* (nt838A) reaction, the Kidd normal (exon 6 nt-1G) reaction, and the Kidd null (exon 6 nt-1A) reaction. Allele-specific amplification of either *JKA* or *JKB* generates a 97-bp product, and allele-specific amplification of the intron 5 product generates a 154-bp product. A 423-bp HGH internal positive control product is detected in each reaction possessing DNA template. Set A: Analysis of a Caucasian *JKA/JKA* homozygote; set B: Analysis of a Caucasian *JKB/JKB* homozygote; set C: Analysis of a Polynesian Jk(a–b–) individual. Fifteen to 20 μL of each amplified PCR product was analyzed by electrophoresis through 2% agarose gels and stained with ethidium bromide.

not available. All Polynesian Jk(a–b–) individuals ($n = 10$) typed were homozygous for a silent *JKB* allele possessing the intron 5 G→A 3′ splice acceptor mutation *(69)*. Irshaid et al. reported the absence of any mutations associated with silent *JK* alleles among 64 Swedish Caucasians and detection of 8 intron 5 heterozygotes among 46 Polynesians *(63)*. Together, these limited results indicate that the silent Polynesian *JK* allele, based on its low frequency, is unlikely to cause false positives when conducting prenatal genotyping of individuals from non-Polynesian groups; however, the risk of potential misinterpretation of genotyping results must be considered, especially because this mutation has been observed as the molecular basis of the Jk(a–b–) phenotype in a single Chinese-American *(68)*. The existence of this rare allele again highlights the importance of establishing concordance between the parental genotypes and serotypes prior to predicting the fetal serotype based on genotyping results.

5. GENOTYPING PLATELET AND NEUTROPHIL ANTIGEN SYSTEMS

Neonatal alloimmune thrombocytopenic purpura (NATP), the platelet counterpart of HDN, occurs with a frequency of approx 1 in 1100 live births *(70)*. Infants affected with NATP are at risk of bleeding and 10–30% of cases develop intracranial hemorrhage (50% *in utero*), which can result in brain damage and lasting neurological sequelae. There are currently 25 different platelet-specific alloantigens and antigens shared with other tissues that have been implicated in causing NATP (Table 1). Maternal antibodies produced against the HPA-1a (PI^{A1}) alloantigen are responsible for the vast majority (85%) of serologically confirmed NATP cases followed by antibodies against HPA-5b (Br^a, 15%). However, unlike for HDN resulting from RHD, antenatal screening of women for HPA-1a is not performed, and there is no treatment equivalent to RhoGAM, although some maternal antenatal treatment regimens, including IV-immunoglobulin and/or steroids have been effective in reducing the incidence of infant intercranial hemorrhage *(71)*. The severity of thrombocytopenia in subsequent children born to a mother with a previous NATP-affected pregnancy is the same or greater and the mortality resulting from bleeding and intracranial hemorrhage is 1–3% in untreated cases, therefore, genotyping of the parents and fetuses in subsequent pregnancies provides important information to the obstetrician. For many years, platelet antigens could only be directly typed using serological methods. Discovery of the HPA-1a/b gene polymorphism by Newman et al. in 1989 *(72)* led to the development of several genotyping methods for all of the various platelet gene polymorphisms, recently reviewed by Hurd et al. *(73)*. All but one of the platelet-specific alloantigens are the result of single-nucleotide polymorphisms (SNPs) in one of only five different platelet glycoprotein genes with 10 of the 18 mutations on the GPIIIa gene (Table 1). Genomic DNA isolated from fetal blood, amniocytes, or chorionic villus samples can be used successfully in ASPCR-based assays, which is currently the most commonly used method for platelet genotyping *(5,73)*. As has been observed in the red cell antigen systems discussed earlier, variant alleles have been reported in the GPIIIa gene that result in discordance between the genotype and platelet phenotypes, but the occurrence of these unusual mutations appears to be rare *(74–77)*.

Neonatal alloimmune neutropenia (NAN) is yet another immune cytopenic disorder involving immune destruction of fetal neutrophils by maternal IgG antibodies targeting paternally

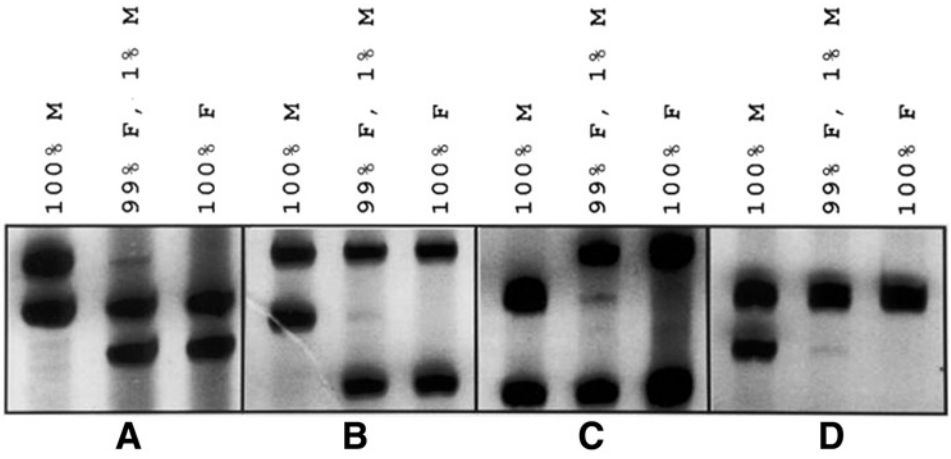


Fig. 4. VNTR analysis of uncultured fetal amniocentesis samples using the D1S80 locus. In sets **A–D** the amplification products from 100% maternal DNA (M), 99% fetal DNA (F) mixed with 1% maternal DNA (M), and 100% fetal DNA (F) are shown. The assay sensitivity is 1% because of the detection of the maternal band in the 99% fetal DNA mixed with 1% maternal DNA lane. Maternal contamination of greater than 1% is not observed in these samples because of the absence of the maternal band in the 100% fetal DNA lane.

inherited neutrophil alloantigens. The incidence of NAN has been reported to be 1 in 500 live births, with infants susceptible to infections following birth *(78)*. Of the neutrophil alloantigens currently identified, the HNA-1a (NA1), HNA-1b (NA2), HNA-1c (SH), and HNA-2a (NB1) antigens are most often implicated in NAN *(78)* and are routinely typed using ASPCR *(8,13,78)*.

6. MATERNAL CONTAMINATION AND THE SENSITIVITY OF GENOTYPING

Maternal contamination of umbilical cord blood, chorionic villus, and amniocentesis samples is a well-documented problem that presents itself to laboratories conducting prenatal genetic testing *(79–82)*. Because analysis by molecular diagnostic methodologies does not distinguish maternal from fetal DNA, maternal contamination presents a risk for misidentification of the fetal genotype and possible misdiagnosis. The reality of this problem is further emphasized when one considers that the cell counts of amniotic fluid samples range from several hundred cells/per milliliter prior to 10 wk gestation to several thousand cells/per milliliter prior to 20 wk gestation. Therefore, the introduction of even small volumes of maternal blood could cause the number of nucleated maternal cells to outnumber the fetal amniocytes, because the average leukocyte count in adults is approx 7×10^6 cells/mL *(82)*. To address this issue, a variable number of tandem repeats (VNTR) analysis of defined sensitivity was conducted on actual fetal samples to define the rate and degree to which maternal contamination can be expected during routine diagnostic testing *(11)*. Experiments were also conducted to determine the extent to which maternal contamination could be tolerated in PCR-based genotyping assays for *RHC/c, RHD, RHE/e*, Kell, HPA-1 (Pl^a), HPA-5 (Br), and FcγRIIIb. In each case, the ability to sensitively detect the possible alloantigen was evaluated, because its presence conveys risk of the immune cytopenic disorder.

The VNTR loci can be used to detect and estimate the degree of maternal contamination within individual fetal samples. The VNTR loci are highly polymorphic and are inherited in a Mendelian fashion, making their analysis useful to confirm (or exclude) the source of template DNA. A total of 56 fetal samples (31 uncultured amniocyte, 6 cultured amniocyte, 18 fetal blood, and 1 chorionic villus) were analyzed. The VNTR loci employed, D1S80 and apoB, were amplified using primers that flank the repeat unit to generate variably sized products. Products were analyzed through polyacrylamide gel electrophoresis and silver staining *(83–85)*. In fetal samples free of maternal contamination, only one of the two maternal bands will be detected in addition to the paternal band. In order to establish the sensitivity at which maternal contamination could be detected in fetal DNA samples, "contaminated" samples were created in cases for which there was adequate material by mixing 1% maternal DNA and 99% fetal DNA. Using this approach, 1% maternal DNA mixed with 99% fetal DNA was always detected in these experiments (Fig. 4) *(11)*. Of the 56 fetal samples analyzed, maternal contamination was only observed in 4 uncultured amniocentesis samples *(11)*. Maternal contamination of less than 1%, based on comparison of band intensities was observed in three of these samples (data not shown). In these cases, the fetuses were found at risk for their respective immune cytopenic disorder. Therefore, the level of maternal contamination present in these samples did not interfere with the detection of the paternal alloallele in the fetus. The fourth sample, which was visibly blood stained, possessed maternal contamination exceeding 50%, based on comparison of band intensities. A portion of this amniotic fluid sample was cultured and later found free of any detectable maternal contamination by VNTR analysis. This is consistent with the recent report of Winsor et al., which describes much higher maternal contamination rates in uncultured (21.4%) vs cultured (0.2%) amniotic fluid samples *(86)*. In this particular case, the cultured and uncultured amniocentesis samples were subjected to HPA-1 genotyping. Both genotypings indicated that the fetus was not at risk for NATP.

In immune cytopenic disorders, a paternal allele encoding an alloantigen is responsible for the incompatibility between the mother and the fetus. Therefore, when identifying fetuses at risk, it is critical that this paternal alloallele be detected. To

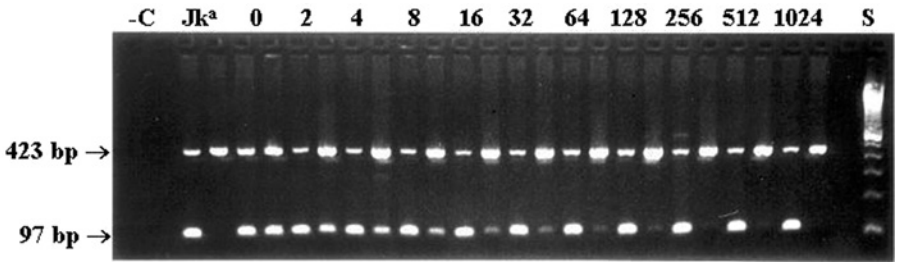


Fig. 5. The sensitivity of ASPCR for detection of the *JKB* allele of the Kidd blood group system. Products of each analysis are run in pairs; the *JKA* reaction is in the first lane of each pair, and the *JKB* reaction is in the second lane of each pair. A 423-bp HGH control product is detected in each reaction possessing, the DNA template. Allele-specific amplification of either *JKA* or *JKB* generates a 97-bp product. –C: ASPCR reactions with no DNA template; Jk^a: ASPCR reactions with DNA template isolated from a Jk^a/Jk^a serotyped individual; 0: ASPCR reactions with DNA template isolated from a Jk^a/Jk^b serotyped individual; 2–1024: ASPCR reactions with DNA template isolated from a Jk^a/Jk^b serotyped individual serially diluted into DNA template isolated from a Jk^a/Jk^a serotyped individual (1 : 2 to 1 : 1024). The last lane is 100-bp DNA ladder size standard (Gibco/BRL, Gaithersburg, MD).

determine the degree of maternal contamination in fetal samples that could be tolerated in ASPCR genotyping assays without the loss of detection of this paternal allele, maternal contamination of fetal samples was mimicked by diluting DNA isolated from a heterozygous individual into DNA isolated from a homozygous individual (undiluted to 1 : 1024) *(86)*. Figure 5 illustrates an example of the results obtained when performing sensitivity studies using the Kidd ASPCR genotyping assay. In this example, DNA isolated from a person heterozygous for the Kidd antigen system (to mimick the fetus) is diluted into DNA isolated from a homozygous individual (to mimick the mother). The K1 allele is detectable to a point where the heterozygote DNA comprises only 0.2% of the template DNA in the reaction. Similar experiments have been conducted on all other ASPCR assays used for prenatal testing and indicate that >90% maternal contamination can be tolerated in these assays without the loss of detection of the paternal allele *(86)*. Such samples would likely be blood stained. Therefore, a large safety margin exists when one considers the frequency and degree to which fetal amniocyte samples are contaminated and the ability of ASPCR assays to detect the offending paternal alloallele. It is advisable that laboratories using other typing methodologies define the sensitivities of the assays/technologies that they are employing. It is also advisable to confirm the fetal origin of DNA samples by VNTR analysis when a fetal sample genotypes compatible with the mother prior to withdrawing any further monitoring of the pregnancy.

7. CONCLUSIONS

The availability of molecular genotyping assays, which can accurately detect the presence of paternal alloalleles, provide an important tool in managing pregnancies at risk for immune cytopenic disorders. Despite their utility and effectiveness, molecular diagnostic methodologies are not free of potential problems. Users must be aware that the presence of a gene does not necessarily correspond to the expression of the encoded protein antigen on the cell surface. A limitation of ASPCR, as well as other approaches, is that new allelic variants might not be detected because they might not possess sequences to which specific primers or probes are directed. Therefore, it is necessary to identify discordance between parental serotypes and genotypes when conducting prenatal genotyping for immune cytopenic disorders. When using fetal genotyping alone, failure to identify inconsistencies between parental serotypes and genotypes can lead to an incorrect assumption of fetal serology. Even with this precaution, although unlikely, it is conceivable that certain variants or combinations of variants might be undetectable. Additionally, the possibility of nonpaternity can negate the effort of establishing concordance between the "paternal" serotype and genotype.

The extreme sensitivity of PCR-based assays is beneficial when attempting to identify antigen incompatibilities between a pregnant woman and her fetus. Even extensive maternal contamination of an amniotic sample, which we have found to be infrequent, is unlikely to interfere with the detection of an incompatible allele in the fetus, because only a few copies of a template DNA are required for its detection. However, it is advisable to confirm the fetal origin of DNA used for genotyping by VNTR analysis before withdrawing further monitoring or therapy in fetuses identified as being not at risk.

In sensitized women, where the father's blood type is unknown or the father is heterozygous for the antigen system in question, establishing the fetal blood type can be accomplished through cordocentesis. However, the risk of pregnancy loss with cordocentesis is greater than the risk associated with amniocentesis, and the ability to serotype the fetal blood sample might be complicated by the presence of maternal antibody. Genotyping, using DNA isolated from amniocytes, is an especially useful tool for identifying fetuses at risk for immune cytopenic disorders. In cases where the father is potentially incompatible with the mother, or his phenotype is unknown, establishing that the fetus lacks the relevant allele and is, therefore, not at risk, obviates the need for expensive and invasive monitoring by amniocentesis throughout the pregnancy. Identifying a fetus as being at risk, on the other hand, allows for appropriate monitoring, early intervention, and improved care.

REFERENCES

1. Liley, A. W. Liquor amnii analysis in management of pregnancy complicated by rhesus immunization. *Am. J. Obstet. Gynecol.* **82:** 1359–1371, 1961.
2. Bowman, J. M. and Pollock, J. M. Transplacental fetal hemorrhage after amniocentesis. *Obstet. Gynecol.* **66:**749–54, 1985.

3. Hall, J. G., Eis, P. S., Law, S. M., et al. Sensitive detection of DNA polymorphisms by the serial invasive signal amplification reaction. *Proc. Natl. Acad. Sci. USA*, **97:**8272–8277, 2000.
4. Kaiser, M. W., Lyamicheva, N., Ma, W., et al. A comparison of eubacterial and archaeal structure-specific 5′–exonucleases. *J. Biol. Chem.* **274:**21,387–21,394, 1999.
5. Skogen, B., Bellissimo, D. B., Hessner, M. J., et al. Rapid determination of platelet alloantigen genotypes by polymerase chain reaction using allele–specific primers. *Transfusion* **34:**955–960, 1994.
6. Pearson, S. L. and Hessner, M. J. A(1,2)BO(1,2) genotyping by multiplexed allele-specific PCR. *Br. J. Haematol.* **100:**229–234, 1998.
7. Hessner, M. J., Dinauer, D. M., Kwiatkowski, R., Neri, B., and Raife, T. J. Age–dependent prevalence of vascular disease–associated polymorphisms among 2689 volunteer blood donors. *Clin. Chem.* **47:**1879–1884, 2001.
8. Hessner, M. J., Shivaram, S. M., Dinauer, D. M., Curtis, B.R., Endean, D. J., and Aster, R. H. Neutrophil antigen (FcgammaRIIIB) SH gene frequencies in six racial groups. *Blood* **93:**1115–1116, 1999.
9. Hessner, M. J., Pircon, R. A., Johnson, S. T., and Luhm, R. A. Prenatal genotyping of the Duffy blood group system by allele–specific polymerase chain reaction. *Prenat. Diagn.* **19:** 41–45, 1999.
10. Hessner, M. J., Pircon, R. A., Johnson, S. T., and Luhm, R. A. Prenatal genotyping of Jk(a) and Jk(b) of the human Kidd blood group system by allele–specific polymerase chain reaction. *Prenat. Diagn.* **18:**1225–1231, 1998.
11. Hessner, M. J., Agostini, T. A., Bellissimo, D. B., Endean, D. J., Pircon, R. A., and Kirschbaum, N. E. The sensitivity of allele-specific polymerase chain reaction can obviate concern of maternal contamination when fetal samples are genotyped for immune cytopenic disorders. *Am. J. Obstet. Gynecol.* **176:**327–333, 1997.
12. Hessner, M. J., McFarland, J. G., and Endean, D. J. Genotyping of KEL1 and KEL2 of the human Kell blood group system by the polymerase chain reaction with sequence-specific primers. *Transfusion* **36:**495–499, 1996.
13. Hessner, M. J., Curtis, B. R., Endean, D. J., and Aster, R. H. Determination of neutrophil antigen gene frequencies in five ethnic groups by polymerase chain reaction with sequence-specific primers. *Transfusion* **36:**895–899, 1996.
14. Ekman, G. C., Billingsly, R., and Hessner, M. J. Rh genotyping: avoiding false–negative and false-positive results among individuals of African ancestry. *Am. J. Hematol.* **69:**34–40, 2002.
15. Cherif-Zahar, B., Mattei, M. G., Le Van Kim, C., Bailly, P., Cartron, J. P., and Colin, Y. Localization of the human Rh blood group gene structure to chromosome region 1p34.3–1p36.1 by in situ hybridization. *Hum. Genet.* **86:**398–400, 1991.
16. Colin, Y., Cherif-Zahar, B., Le Van Kim, C., Raynal,V., Van Huffel, V., and Cartron, J. P. Genetic basis of the RhD-positive and RhD–negative blood group polymorphism as determined by Southern analysis. *Blood* **78:**2747–2752, 1991.
17. Le van Kim, C., Mouro, I., Cherif–Zahar, B., et al. Molecular cloning and primary structure of the human blood group RhD polypeptide. *Proc. Natl. Acad. Sci. USA*, **89:**10,925–10,929, 1992.
18. Arce, M. A., Thompson, E. S., Wagner, S., Coyne, K. E., Ferdman, B. A., and Lublin, D. M. Molecular cloning of RhD cDNA derived from a gene present in RhD-positive, but not RhD-negative individuals. *Blood* **82:**651–655, 1993.
19. Mouro, I., Colin, Y., Cherif-Zahar, B., Cartron, J. P., and Le Van Kim, C. Molecular genetic basis of the human Rhesus blood group system. *Nat. Genet.* **5:**62–65, 1993.
20. Siegel, D. L. The human immune response to red blood cell antigens as revealed by repertoire cloning. *Immunol. Res.* **17:**239–251, 1998.
21. Rossie, E. C., Simon, T. L., and Moss, G. S. *Principles of Transfusion Medicine*. 2nd ed Williams and Wilkins, Baltimore, MD, 1996.
22. Beckers, E. A., Faas, B. H., Simsek, S. et al. The genetic basis of a new partial D antigen: DDBT. *Br. J. Haematol.* **93:**720–727, 1996.
23. Bennett, P. R., Le Van Kim, C., Colin, Y., et al. Prenatal determination of fetal RhD type by DNA amplification. *N. Engl. J. Med.* **329:** 607–610, 1993.
24. Wolter, L. C., Hyland, C. A., and Saul, A. Rhesus D genotyping using polymerase chain reaction. *Blood* **82**:1682–1683, 1993.
25. Simsek, S., Faas, B. H., Bleeker, P. M., et al. Rapid Rh D genotyping by polymerase chain reaction-based amplification of DNA. *Blood* **85:** 2975–2980, 1995.
26. Mouro, I., Le Van Kim, C., Rouillac, C., et al. Rearrangements of the blood group RhD gene associated with the DVI category phenotype. *Blood* **83:**1129–1135, 1994.
27. Singleton, B. K., Green, C. A., Avent, N. D., et al. The presence of an RHD pseudogene containing a 37 base pair duplication and a nonsense mutation in africans with the Rh D-negative blood group phenotype. *Blood* **95:**12–18, 2000.
28. Aubin, J. T., Le Van Kim, C., Mouro, I., et al. Specificity and sensitivity of RHD genotyping methods by PCR-based DNA amplification. *Br. J. Haematol.* **98:**356–364, 1997.
29. Okuda, H., Kawano, M., Iwamoto, S., et al. The RHD gene is highly detectable in RhD-negative Japanese donors. *J. Clin. Invest.* **100:** 373–379, 1997.
30. Daniels, G., Green, C., and Smart, E. Differences between RhD-negative Africans and RhD-negative Europeans. *Lancet* **350:** 862–863, 1997.
31. Race, R. R., and Sanger, R. *Blood Groups in Man,* 6th ed. Blackwell Scientific, Oxford, 1975.
32. Chiu, R. W., Murphy, M. F., Fidler, C., Zee, B. C., Wainscoat, J. S., and Lo, Y. M. Determination of RhD zygosity: comparison of a double amplification refractory mutation system approach and a multiplex real–time quantitative PCR approach. *Clin. Chem.* **47:**667–672, 2001.
33. Walknowska, J., Conte, F. A., and Grumbach, M. M. Practical and theoretical implications of fetal–maternal lymphocyte transfer. *Lancet* **1:**1119–1122, 1969.
34. Lo, Y. M., Patel, P., Wainscoat, J. S., Sampietro, M., Gillmer, M. D., and Fleming, K. A. Prenatal sex determination by DNA amplification from maternal peripheral blood. *Lancet* **2:**1363–1365, 1989.
35. Bianchi, D. W. Fetal cells in the maternal circulation: feasibility for prenatal diagnosis. *Br. J. Haematol.* **105:**574–583, 1999.
36. Strickland, S. and Richards, W. G. Invasion of the trophoblasts. *Cell* **71:**355–357, 1992.
37. Lo, Y. M., Corbetta, N., Chamberlain, P. F., et al. Presence of fetal DNA in maternal plasma and serum. *Lancet* **350:**485–487, 1997.
38. Lo, Y. M. Fetal DNA in maternal plasma: biology and diagnostic applications. *Clin. Chem.* **46:**1903–1906, 2000.
39. Bianchi, D. W., Zickwolf, G. K., Weil, G. J., Sylvester, S., and DeMaria, M. A. Male fetal progenitor cells persist in maternal blood for as long as 27 years postpartum. *Proc. Natl. Acad. Sci. USA*, **93:**705–708, 1996.
40. Lo, Y. M., Tein, M. S., Lau, T. K., et al. Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. *Am. J. Hum. Genet.* **62:**768–775, 1998.
41. Lo, Y. M., Zhang, J., Leung, T. N., Lau, T. K., Chang, A. M., and Hjelm, N.M. Rapid clearance of fetal DNA from maternal plasma. *Am. J. Hum. Genet.* **64:**218–224, 1999.
42. Lo, Y. M., Hjelm, N. M., Fidler, C., et al. Prenatal diagnosis of fetal RhD status by molecular analysis of maternal plasma. *N. Engl. J. Med.* 339:1734–1738, 1998.
43. Bischoff, F. Z., Nguyen, D. D., Marquez-Do, D., Moise, K. J., Jr., Simpson, J. L., and Elias, S. Noninvasive determination of fetal RhD status using fetal DNA in maternal serum and PCR. *J. Soc. Gynecol. Invest.* **6:**64–69, 1999.
44. Faas, B. H., Beuling, E. A., Christiaens, G. C., von dem Borne, A. E., and van der Schoot, C. E. Detection of fetal RHD-specific sequences in maternal plasma. *Lancet* **352:**1196, 1998.
45. Zhong, X. Y., Holzgreve, W., and Hahn, S. Detection of fetal Rhesus D and sex using fetal DNA from maternal plasma by multiplex polymerase chain reaction. *Bjog* **107:**766–769, 2000.
46. Cherif-Zahar, B., Bloy, C., Le Van Kim, C., et al. Molecular cloning and protein structure of a human blood group Rh polypeptide. *Proc. Natl. Acad. Sci. USA.* **87:**6243–6247, 1990.
47. Simsek, S., de Jong, C. A., Cuijpers, H.T., et al. Sequence analysis of cDNA derived from reticulocyte mRNAs coding for Rh

polypeptides and demonstration of E/e and C/c polymorphisms. *Vox Sang* **67:**203–209, 1994.
48. Hyland, C. A., Wolter, L. C., Liew, Y. W., and Saul, A. A southern analysis of Rh blood group genes: association between restriction fragment length polymorphism patterns and Rh serotypes. *Blood* **83:**566–572, 1994.
49. Wolter, L. C., Hyland, C. A., and Saul, A. Refining the DNA polymorphisms that associate with the rhesus c phenotype. *Blood* **84:**985–986, 1994.
50. Le Van Kim, C., Mouro, I., Brossard, Y., Chavinie, J., Cartron, J. P., and Colin, Y. PCR-based determination of Rhc and RhE status of fetuses at risk of Rhc and RhE haemolytic disease. *Br. J. Haematol.* **88:**193–195, 1994.
51. Faas, B. H., Simsek, S., Bleeker, P. M., et al. Rh E/e genotyping by allele–specific primer amplification. *Blood* **85:**829–832, 1995.
52. Flegel, W. A., Wagner, F. F., Muller, T. H., and Gassner, C. Rh phenotype prediction by DNA typing and its application to practice. *Transfus. Med.* **8:**281–302, 1998.
53. Gassner, C., Schmarda, A., Kilga-Nogler, S., et al. RHD/CE typing by polymerase chain reaction using sequence-specific primers. *Transfusion* **37:**1020–1026, 1997.
54. Avent, N. D., Daniels, G., Martin, P. G., Green, C. A., Finning, K. M., and Warner, K. M. Molecular investigation of the Rh C/c polymorphism. *Transfus. Med.* **7(Suppl. 1):**18, 1997.
55. Tax, M. G., van der Schoot, C. E., van Doorn, R., Douglas-Berger, L., van Rhenen, D. J., and Maaskant-vanWijk, P. A. RHC and RHc genotyping in different ethnic groups. *Transfusion* **42:**634–644, 2002.
56. Marsh, W. L. and Redman, C. M. The Kell blood group system: a review. *Transfusion* **30:**158–167, 1990.
57. Redman, C. M., Marsh, W. L., Mueller, K. A., Avellino, G. P., and Johnson, C. L. Isolation of Kell-active protein from the red cell membrane. *Transfusion* **24:**176–178, 1984.
58. Lee, S., Wu, X., Reid, M., Zelinski, T., and Redman, C. Molecular basis of the Kell (K1) phenotype. *Blood* **85:**912–916, 1995.
59. Lee, S., Zambas, E., Green, E. D., and Redman, C. Organization of the gene encoding the human Kell blood group protein. *Blood* **85:**1364–1370, 1995.
60. Issett, P. D. *Applied Blood Group Serology*, 5th ed. Montgomery Scientific Miami, FL, 1989.
61. Bern, M. M. and Frigoletto, F. D. *Hematologic Disorders in Maternal-Fetal Medicine*. Wiley-Liss, New York, 1990.
62. Lee, S., Bennett, P. R., Overton, T., Warwick, R., Wu, X., and Redman, C.M. Prenatal diagnosis of Kell blood group genotypes: KEL1 and KEL2. *Am. J. Obstet. Gynecol.* **175:**455–459, 1996.
63. Irshaid, N. M., Henry, S. M., and Olsson, M. L. Genomic characterization of the Kidd blood group gene:different molecular basis of the Jk(a–b–) phenotype in Polynesians and Finns. *Transfusion* **40:**69–74, 2000.
64. Olives, B., Neau, P., Bailly, P., et al. Cloning and functional expression of a urea transporter from human bone marrow cells. *J. Biol. Chem.* **269:**31,649–31,649, 1994.
65. Olives, B., Mattei, M. G., Huet, M., et al. Kidd blood group and urea transport function of human erythrocytes are carried by the same protein. *J. Biol. Chem.* **270:**15,607–15,610, 1995.
66. Olives, B., Martial, S., Mattei, M.G., et al. Molecular characterization of a new urea transporter in the human kidney. *FEBS Lett.* **386:**156–160, 1996.
67. Olives, B., Merriman, M., Bailly, P., et al. The molecular basis of the Kidd blood group polymorphism and its lack of association with type 1 diabetes susceptibility. *Hum. Mol. Genet.* **6:**1017–1020, 1997.
68. Lucien, N., Sidoux-Walter, F., Olives, B., et al. Characterization of the gene encoding the human Kidd blood group/urea transporter protein. Evidence for splice site mutations in Jknull individuals. *J. Biol. Chem.* **273:**12,973–12,980, 1998.
69. Ekman, G. C. and Hessner, M. J. Screening of six racial groups for the intron 5 G→A 3′ splice acceptor mutation responsible for the polynesian kidd (a–b–) phenotype: the null mutation is not always associated with the JKB allele. *Transfusion* **40:**888–889, 2000.
70. Williamson, L. M., Hackett, G., Rennie, J., et al. The natural history of fetomaternal alloimmunization to the platelet-specific antigen HPA-1a (PlA1, Zwa) as determined by antenatal screening. *Blood* **92:**2280–2287, 1998.
71. Bussel, J. B., Berkowitz, R. L., Lynch, L., et al. Antenatal management of alloimmune thrombocytopenia with intravenous gammaglobulin: a randomized trial of the addition of low-dose steroid to intravenous gamma-globulin. *Am. J. Obstet. Gynecol.* **174:**1414–1423, 1996.
72. Newman, P. J., Derbes, R. S., and Aster, R. H. The human platelet alloantigens, PlA1 and PlA2, are associated with a leucine33/proline33 amino acid polymorphism in membrane glycoprotein IIIa, and are distinguishable by DNA typing. *J. Clin. Invest.* **83:**1778–1781, 1989.
73. Hurd, C. M., Cavanagh, G., Schuh, A., Ouwehand, W. H., and Metcalfe, P. Genotyping for platelet-specific antigens: techniques for the detection of single nucleotide polymorphisms. *Vox Sang* **83:**1–12, 2002.
74. Kjaer, K. M., Jaegtvik, S., Husebekk, A., and Skogen, B. Human platelet antigen 1 (HPA 1) genotyping with 5′ nuclease assay and sequence–specific primers reveals a single nucleotide deletion in intron 2 of the HPA 1a allele of platelet glycoprotein IIIa. *Br. J. Haematol.* **117:**405–408, 2002.
75. Skogen, B., Wang, R., McFarland, J. G., and Newman, P. J. A dinucleotide deletion in exon 4 of the PlA2 allelic form of glycoprotein IIIa: implications for the correlation of serologic versus genotypic analysis of human platelet alloantigens. *Blood* **88:**3831–3836, 1996.
76. Perichon, B., Clemenceau, S., Romand, R., Elion, J., Kaplan, C., and Krishnamoorthy, R. An additional HpaII polymorphism in exon 2 of the human platelet membrane glycoprotein IIIa gene. *Hum. Genet.* **93:**353–354, 1994.
77. Unkelbach, K., Kalb, R., Breitfeld, C., Santoso, S., Kiefel, V., and Mueller-Eckhardt, C. New polymorphism on platelet glycoprotein IIIa gene recognized by endonuclease Msp I: implications for PlA typing by allele-specific restriction analysis. *Transfusion* **34:**592–595, 1994.
78. Dale, D. C. Immune and idiopathic neutropenia. *Curr. Opin. Hematol.* **5:**33–36, 1998.
79. Denomme, G. A., Waye, J. S., Burrows, R. F., et al. The prenatal identification of fetal compatibility in neonatal alloimmune thrombocytopenia using amniotic fluid and variable number of tandem repeat (VNTR) analysis. *Br. J. Haematol.* **91:**742–746, 1995.
80. Ganshirt-Ahlert, D., Pohlschmidt, M., Gal, A., Horst, J., Miny, P., and Holzgreve, W. Transabdominal placental biopsy in the second and third trimesters of pregnancy: what is the risk of maternal contamination in DNA diagnosis? *Obstet. Gynecol.* **75:**320–323, 1990.
81. Rebello, M. T., Hackett, G., Smith, J., et al. Extraction of DNA from amniotic fluid cells for the early prenatal diagnosis of genetic disease. *Prenat. Diagn.* **11:**41–46, 1991.
82. Rebello, M. T., Abas, A., Nicolaides, K., and Coleman, D. V. Maternal contamination of amniotic fluid demonstrated by DNA analysis. *Prenat. Diagn.* **14:**109–112, 1994.
83. Budowle, B., Chakraborty, R., Giusti, A. M., Eisenberg, A. J., and Allen, R. C. Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. *Am. J. Hum. Genet.* **48:**137–144, 1991.
84. Boerwinkle, E., Xiong, W. J., Fourest, E., and Chan, L. Rapid typing of tandemly repeated hypervariable loci by the polymerase chain reaction: application to the apolipoprotein B 3′ hypervariable region. *Proc. Natl. Acad. Sci. USA* **86:**212–216, 1989.
85. Bassam, B. J., Caetano-Anolles, G., and Gresshoff, P. M. Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal. Biochem.* **196:**80–83, 1991.
86. Winsor, E. J., Silver, M. P., Theve, R., Wright, M., and Ward, B. E. Maternal cell contamination in uncultured amniotic fluid. *Prenat. Diagn.* **16:**49–54, 1996.

29 Personalized Medicine

KAREN MCCULLOUGH

1. INTRODUCTION

As many of us can attest from personal experience, not everyone has the same response to any given medication. Some of us might find that a specific medication improves our symptoms, whereas others find that it does not. Some of us find that a given medication improves our symptoms, but we choose not to use it because it has untoward effects. The bases for such varied response to pharmacotherapies are many. Age, organ function, concomitant medications, gender, diet, and other factors all influence how individuals respond to drugs. However, genetic variation, an additional contributor to drug response, is increasingly recognized as a critical determinant of the effectiveness and safety/tolerability of pharmacological agents.

The human genome is composed of an estimated 3 billion basepairs. Of these, roughly 0.01%, or 3 million, are variable. Variability of genetic sequence takes a variety of forms, including insertions and deletions of nucleotides, but is most commonly found as single-nucleotide polymorphisms (SNPs). SNPs are distinguished from mutations based on their frequency of occurrence in the population; if the variation occurs in more than 1% of the population, it is described as a SNP, whereas a variation occurring in less 1% of the population is referred to as mutation. Many SNPs have no impact on gene function. However, some occur in coding regions and result in truncation or altered formation of the resultant gene product. SNPs occurring outside of coding regions, particularly in promoters and splice sites, can also have profound impacts on the amount or function of the resultant gene product. The study of the effects of genetic variation on drug response is known as pharmacogenetics *(1–5)*.

Once a drug is administered, it is absorbed and distributed to its site of action, where it interacts with targets (receptors, enzymes). The drug undergoes metabolism and is then excreted. Genes that regulate each of these processes could potentially be polymorphic. By altering the amount of drug accessible to the drug target, or the nature and effects of interaction between the drug and drug target, such genetic variation can have significant clinical consequences. Clinical observations of inherited differences in drug effects were first documented some 40 yr ago when it was reported that impairment of the metabolism of a muscle relaxant by pseudocholinesterase was a familial trait *(6)*. Soon after this finding was reported, it was discovered that a common genetic defect in the drug-metabolizing enzyme *N*-acetyltransferase could impact the plasma concentrations of a variety of drugs, including the antituberculosis agent isoniazid *(7)*. By the early 1960s, the impact of inheritance on drug response was well recognized, and its study was termed Pharmacogenetics *(8–11)*. Although the field has been in existence for some time, it did not become a primary area of research until the 1990s. As a result of the Human Genome Project and the advent of genomics-era technologies, pharmacogenetics has advanced tremendously.

2. DRUG METABOLISM AND DISTRIBUTION

As described earlier, the field of pharmacogenetics began with examinations of polymorphisms in drug-metabolizing enzymes. Although the field has expanded well beyond drug-metabolizing enzymes, this aspect of pharmacogenetics continues to be the most thoroughly developed. In large part this is because these traits are primarily monogenic (polymorphisms in one gene are responsible for the observed phenotype) and highly penetrant (likely to influence phenotype).

The metabolism of any given drug might involve one or many enzymes. The metabolic process usually involves transforming a lipophilic molecule into one that is hydrophilic. To be absorbed and gain access to their intended target, drugs must be fairly lipid soluble. However, lipophilic molecules cannot easily be excreted from the body. The metabolic reactions involved in the biotransformation of drugs are divided into two categories: phase I and phase II reactions. Phase I reactions oxidize, reduce, or hydrolyze drugs *(12,13)*. Other metabolic pathways involve conjugation reactions, including acetylation, glucuronidation, sulfation, and methylation *(13)*. The metabolism of some drugs involves phase I followed by phase II reactions. However, this nomenclature is primarily historical and different sequences of metabolic processes are found; phase II reactions can precede phase I reactions, and certain compounds are conjugated without first undergoing phase I biotransformation.

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

2.1. PHASE I METABOLISM The cytochrome P450 (CYP) enzymes are a superfamily of microsomal drug-metabolizing enzymes that are primarily responsible for catalyzing phase I reactions *(12,14)*. Three CYP families, CYP1, CYP2 and CYP3, are responsible for drug metabolism *(14)*. Importantly, the genes encoding these enzymes are all polymorphic, and some of these polymorphisms have clearly documented functional importance *(15–17)*. The impact of genetic variation on drug metabolism is illustrated here by *CYP2D6*, although analogous variations can be found in *CYP2C9* and *CYP2C19*. *CYP2D6* is an extensively studied and highly variable isoform of the cytochrome P450s that is responsible for the metabolism of many clinically important drugs *(18,19)*. At least 15 variants of *CYP2D6* have been identified that produce enzyme that is either nonfunctional or has altered function *(20)*. Defective alleles occur at extremely variable frequencies in racially diverse populations *(21,22)*. However, two null alleles are the most commonly occurring variants of *CYP2D6* associated with diminished CYP2D6 enzymatic activity *(23,24)*. In addition to variations that diminish or eliminate the activity of the *CYP2D6* gene product, polymorphisms have also been identified that increase the activity of the enzyme. Enhanced CYP2D6 activity results from two main mechanisms. Gene duplication has been identified in some families, with the copy number of *CYP2D6* ranging from 2 to 13 *(25–27)*. In addition, transcription-enhancing polymorphisms have been identified in the promoter region of *CYP2D6 (20)*.

Thus, for *CYP2D6*, four phenotypes are possible: poor metabolizers (PMs), who have greatly diminished activity of the enzyme; intermediate metabolizers (IMs), who carry one deficient and one normal allele; extensive metabolizers (EMs) who carry two normal copies of the gene; and ultrametabolizers (UM), who have heightened CYP2D6 activity. The effect of *CYP2D6* variations is that drug metabolism might vary as much as 1000-fold between PMs and UMs *(15,20)*. The clinical impact of this can be enormous. All drugs have therapeutic windows, meaning that exposure within a certain range is likely to produce a desired, beneficial effect; exposure below this threshold will not be effective, and higher exposure results in toxicity, potentially the result of interactions with unintended targets *(28)*. Dosing recommendations for drugs are developed for the general population, most of whom are EMs. However, for the roughly 7% of the population who are PMs, the recommended dose is far too high and will likely be toxic *(29)*. Conversely, for the roughly 5.5% of the population who are UMs, the drug will be metabolized so quickly that exposure will be too low to be effective. In the case of life-threatening diseases, where rapid intervention in the disease process is critical, lack of efficacy can be just as devastating as the induction of a severe adverse drug reaction.

Characterization of the effects of *CYP2D6* variation on metabolism of drugs during their development is becoming more common. For some drugs, strategies for modulating dose to accommodate genotypes are recommended on the drug's label. Strattera® (atomoxetine HCl) is a selective norepinephrine reuptake inhibitor approved in the United States for the treatment of attention-deficit/hyperactivity disorder (ADHD). Atomoxetine is metabolized primarily through the action of CYP2D6 *(30)*. PMs have elevated plasma concentrations of atomoxetine *(31,32)*, which does not result in toxicity but can become a problem if atomoxetine is coadministered with other drugs that inhibit CYP2D6. Examples of such drugs include fluoxetine, paroxetine, and quinidine. The manufacturers of atomoxetine recommend identifying a patient's *CYP2D6* genotype before administering it with a CYP2D6 inhibitor.

2.2. PHASE II METABOLISM Examples of the impact of genetic variation on drug metabolism are not limited to phase I reactions. There are more than 30 families of enzymes that mediate phase II metabolic reactions, and a number of these have clinically relevant genetic variants *(5,29)*. One of the most widely recognized examples of a polymorphic phase II enzyme is thiopurine *S*-methyltransferase (TPMT). TPMT is responsible, in part, for the metabolism of thiopurine drugs *(33,34)*, including 6-mercaptopurine and 6-thioguanine, which are used in the treatment of lymphoblastic leukemias *(35)*, and azathioprine, a widely used immunosuppressant given to patients with autoimmune conditions, as well as transplant recipients *(36,37)*. These drugs are converted to active metabolites in vivo that are highly toxic and have narrow therapeutic windows *(38,39)*. TPMT transfers a methyl group to thiopurines and reduces their bioavailability for conversion into cytotoxic metabolites *(40,41)*. In the absence of *TPMT*, toxic metabolites accumulate in hematopoietic tissues and can lead to severe and potentially fatal hematological toxicities *(42)*. TPMT is coded for by a highly polymorphic gene; to date at least 10 variations in *TPMT* have been associated with low TPMT activity *(40)*. In the United States about 10% of whites and African-Americans are heterozygous for a defective form of *TPMT* and have intermediate TPMT activity as a result. In addition, roughly 0.3% of these populations are homozygous for a defective *TMPT* allele and demonstrate essentially no TPMT activity *(43)*. For such individuals, inactivation of thiopurines is so compromised, severe and potentially life-threatening toxicity is likely to result if normal dosages are administered. To compensate for reduced TPMT activity, it is recommended that patients heterozygous for defective *TPMT* alleles be given 65% of the standard thiopurine dosage, whereas homozygous patients should receive 6–10% of the standard dosage *(43)*. As long as dosages are adjusted, thiopurines can be very effective in treating disease in individuals with intermediate or low TPMT activity, and detection of the presence of defective *TPMT* alleles can be accomplished using simple tests *(44,45)*. Indeed, DNA tests for inactive TPMT alleles were among the first pharmacogenetic tests implemented in clinical practice *(5,44)*.

2.3. TRANSPORTERS In addition to drug-metabolizing enzymes, transporters are critical determinants of drug exposure. Transporters are transmembrane proteins that control the efflux of drugs from cells and thus impact the absorption, distribution, and excretion of many medications *(46–48)*. Members of the adenosine triphosphate (ATP)-binding cassette family of transporters, the so-called ABCB family, are among the most extensively studied transporters involved in drug disposition *(49)*. The first ABCB transporter to be cloned, P-glycoprotein (Pgp), has been extensively characterized. The principle function of Pgp is the ATP-dependent efflux of substrates from cells. The protein has broad substrate specificity, transporting

diverse molecules, including anticancer drugs, cardiac glycosides, immunosuppressive agents, and glucocorticoids, and is expressed in a wide variety of tissues *(49–51)*. Given the wide tissue distribution of the drug and its activity at critical barrier sites, including the blood–brain and blood–testes barrier and the lower gastrointestinal tract, Pgp is well positioned to prevent the entry of toxic compounds into the body or into specific compartments of the body *(52,53)*.

P-glycoprotein is coded for in humans by *ABCB1*, also known as *MDR1*, a highly polymorphic gene. Multiple SNPs in *ABCB1* have been demonstrated to impact drug therapy *(3,28)*. These SNPs have been associated with altered exposure to and/or effect of a variety of drugs, including digoxin *(54)*, the antihistamine fexofenadine *(55,56)*, and the human immunodeficiency virus (HIV) protease inhibitors nelfinavir and efavirenz *(57)*.

3. DRUG TARGETS

Clearly, the extent of exposure to a pharmacological agent has a large impact on clinical outcomes; low exposure compromises efficacy, whereas high exposures result in toxicity. However, there are many examples of varied response to drugs in which the clinical outcomes do not correlate with blood or tissue levels of the drug and, therefore, do not arise from polymorphisms in genes controlling the absorption, distribution, or metabolism of the drug. In such instances, genetic variation might still contribute to varied drug response because polymorphisms can also occur in drug targets, as well as in downstream components of the target's signaling pathway. For a number of reasons, defining associations between drug targets, and drug response has been much more challenging and has seen fewer successes than the pharmacogenetics of drug-metabolizing enzymes.

Complex common disorders arise from the interplay of many factors, including multiple gene–gene and gene–environment interactions, making the role of variation in any given gene difficult to discern. Indeed, the complex common diseases that are the main contributors to society's disease burden are fundamentally different from classical, monogenic, "Mendelian" diseases. Monogenic disorders like cystic fibrosis arise from a defect in a single gene, and the presence of specific defects in the gene results in the expression of the disease phenotype . Conversely, a gene defect associated with complex common diseases is only one of many factors associated with the development of the disease, and although its presence increases the probability of the disease arising, it does not determine the phenotype. In addition, complex diseases are heterogeneous, with related but distinct etiologies. For example, an individual with an inflammatory disease like asthma might, among other gene defects, have a defect in the tumor necrosis factor (TNF) promoter resulting in exaggerated expression of TNF-α. Another asthmatic may share gene defects with the first, but harbor wild-type *TNF* alleles and express normal levels of TNF-α.

Although tremendous progress has been made in the past decade, our molecular understanding of the pathogenesis of common diseases is limited. Because we cannot define subgroups according the molecular etiology of the disease, complex common diseases are defined according to general phenotypic parameters. As a result, studies that investigate the association between genetic polymorphisms and drug response are confounded by allelic heterogeneity of the study population and often produce results that cannot be reproduced. A recent meta-analysis of 370 pharmacogenetic studies found that results from various studies frequently could not be confirmed in subsequent investigations *(58)*. A number of factors likely contribute to this, including diversity among study populations *(3,58)*.

Although this aspect of the field of pharmacogenetics is still in its infancy, there have been success stories involving the polymorphic drug targets and drug response. One example of variant drug targets influencing response involves the β_2-adrenergic receptor and the treatment of asthma. One of the mainstays of asthma treatment is the activation of the β_2-adrenergic receptor by specific agonists, which leads to the relaxation of bronchial smooth muscles and, consequently, bronchodilation *(59)*. The β_2-adrenergic receptor is coded for by *ADRB2*, in which multiple functional polymorphisms have been identified *(60)*. Several variants of *ADRB2* have been associated with altered response to β_2-agonist treatment. Individuals homozygous for the glutamate allele at position 27 of the encoded protein have been found to have higher maximal venodilation in response to isoproternol than those with the glutamine allele *(28,61)*. Individuals carrying two copies of a variant allele that codes for glycine instead of arginine in position 16 have been shown to have fivefold reduced response to β_2 adrenergic agonists. Individuals heterozygous for the variant allele had a threefold reduced response *(62,63)*.

4. PHARMACOGENETICS IN DRUG DEVELOPMENT

Pharmacogenetics has without doubt become a primary field of research, with extensive involvement of both academic labs and industry. Pharmaceutical companies have invested heavily in genomics-era technologies, and many routinely include pharmacogenetics in their clinical studies *(64)*. Some companies make strategic decisions to investigate pharmacogenetics in specific phases of drug development, but increasingly companies are electing to include pharmacogenetics in all phases of their clinical programs. In a typical clinical study conducted to support the development of a new drug, patients are asked if they would like to consent to participation in an exploratory pharmacogenetic substudy. Consent by patients to be included in this type of investigation is independent from their consent to participate in the trial overall. This is a critical point, as there is considerable unease in the general population about genetics and genetic discrimination *(65–67)*. Following isolation of the subjects' DNA, associations are made between polymorphisms, which are usually SNPs but can be other types of variation as well, and clinical outcomes. The outcomes investigated could be a pharmacokinetic parameter (blood level of the drug), an efficacy measure, or the occurrence of an adverse event.

There are two general approaches to conducting a pharmacogenetics study. The most common is the candidate gene approach, in which specific SNPs in specific genes are genotyped and associated with clinical outcomes. Genes that are considered likely to have an impact on drug response are

selected for genotyping of their polymorphic loci. Selection of genes is based on existing knowledge of the mechanism of action of the medication (the drug target as well as other genes in the target pathway) and genes that are known to determine the metabolism and distribution of the drug. Selection of SNPs is based on the impact of the polymorphism on gene function and the frequency of the SNP in the population of interest. There is an ever-expanding body of literature reporting associations between specific polymorphisms and drug response; these can guide the selection of polymorphisms to examine.

The advantage of the candidate gene approach is that it is cost-effective and amenable to high throughput. Analysis of associations between the clinical parameter of interest and the genotypes is manageable as well. The disadvantage is that it is somewhat of a "needle in a haystack" approach. When studying complex diseases in which many genes could influence drug response, it is easy to select the wrong candidate genes for study. The alternative approach is a whole-genome scan, in which SNPs are assayed systematically across the entire genome. There are significant disadvantages associated with this approach as well. Analysis of this data is much more labor-intensive and most companies lack software to automate the process or reduce its complexity. In addition, whole-genome scans are very expensive. Because costs are prohibitive, whole-genome scans are the exception, not the rule, in drug development programs.

Another critical reason why a pharmacogenetic study might fail to find a genetic explanation for a clinical parameter is that alteration in the gene product (e.g., posttranslational modifications) might not be reflected at the gene level. Gene expression arrays (so-called "transcriptomics") and proteomics approaches are also being developed. Conventionally, studies of the entire transcriptome or proteome as they relate to drug effects are referred to as "pharmacogenomics" because they investigate drug effects in the context of the entire genome. "Pharmacogenetics," on the other hand, considers the effects of only a subset of genes on drug action. (Whole-genome scans can also be considered pharmacogenomics.) Application of techniques to profile the entire proteome or transcriptome in clinical drug development is limited by the fact that sample procurement must be minimally invasive. Usually, it is only reasonable to collect blood from study participants. Unless the drug target is a blood cell, gene expression in the blood is often not informative of events occurring in other tissues. Oncology studies are an important exception, because biopsies can be obtained and profiled more readily.

5. TRANSLATING PHARMACOGENETICS TO THE CLINIC

The practice of clinical medicine relies heavily on trial and error; when multiple drugs exist for the treatment of a disease, the physician's choice as to which one will work best for a given patient is extremely subjective. If that drug is not effective, another one is tried. This process is ineffective. The response rates in treatment of common, debilitating diseases like Alzheimer's, cardiac dysrhythmias, depression, osteoporosis, schizophrenia, and rheumatoid arthritis are in the range of 30–60% *(68)*. In view of this relatively low frequency of responders and the high cost and serious consequences of adverse drug reactions, better tools for predicting drug response are needed.

Pharmacogenetics has been investigated for more than 50 yr, and modern-day drug development programs frequently incorporate pharmacogenetic and/or pharmacogenomic investigations, so why is it that genetics and genomics are so rarely applied in clinical practice? First, pharmacogenetics changes the paradigm of prescribing medicine. There is wide acceptance in the medical community of the "trial-and-error" approach to selecting drugs and then selecting dosages based on a default average *(69)*. Second, diagnostic pharmacogenetic tests have to be simple with easily interpreted results. Many clinicians lack expertise in molecular biology and genetics and it is unreasonable to presume that they will undergo significant training in order to prescribe a particular medicine. However, there is reason to believe that acceptance of pharmacogenetic tests is increasing, particularly in specialized fields of medicine.

Pharmacogenetics has been applied to a greater extent in oncology than in other fields of medicine. As described earlier, genetic tests of TPMT are conducted before treatment of pediatric leukemias are initiated with thiopurines. Furthermore, genomics-era advances have been employed in the development of anticancer drugs to exploit genetic and molecular differences between normal and neoplastic cells. Some of these, including Herceptin™, require a simple test to determine whether a given drug is likely to be effective. Herceptin (trastuzumab) is a monoclonal antibody targeted to erbB2, a receptor tyrosine kinase encoded by the *HER2* oncogene, which is amplified and overexpressed in about 30% of breast cancers *(70)*. Use of trastuzumab is only indicated if breast tumors are *HER2* positive, and two types of test are most commonly used to make that assessment. The first test described was an immunohistochemical assay *(70)*, which became the basis for commercially available diagnostics like the Hercept Test (Dako Corp, Carpinteria, CA). Other measures of suitability for trastuzumab treatment rely on fluorescence *in situ* hybridization (FISH) to assess *HER2* gene amplification. Both types of test are widely used to discern *HER2* status among breast cancer patients *(71,72)*. There is widespread acceptance of conducting the diagnostic test before initiating trastuzumab therapy, although it remains to be seen if pharmacogenetic tests can become routine in general medical practices.

6. CONCLUSION

The promise of pharmacogenetics is that "trial-and-error" practices in selecting pharmacotherapies will be replaced by informed tailoring of treatments to the genetic makeup of individual patients. In its full realization, pharmacogenetics would revolutionize the practice of medicine. The premise of pharmacogenetics is clearly not revolutionary, but only in rare cases have pharmacogenetic practices been routinely adopted. Although it is not reasonable to presume that pharmacogenetics will be capable of individualizing medicine in the near future, it is realistic to expect that molecular genetics will continue, in an incremental fashion, to improve our ability to diagnose and treat illnesses with the right drug at the right dose. Step by step we get closer to the goal of personalized medicine.

REFERENCES

1. Kalow, W. Pharmacogenetics and personalised medicine. *Fundam. Clin. Pharmacol.* **16(5):**337–342, 2002.
2. Lindpaintner, K. Pharmacogenetics and the future of medical practice. *J. Mol. Med.* **81(3):**141–153, 2003.
3. Meisel, C., Gerloff, T., Kirchheiner, J., et al. Implications of pharmacogenetics for individualizing drug treatment and for study design. *J. Mol. Med*. **81(3):**154–167, 2003.
4. Prows, C. A. and Prows, D. R. Medication selection by genotype: how genetics is changing drug prescribing and efficacy. *Am. J. Nurs.* **104(5):**60–70, 2004.
5. Weinshilboum, R. Inheritance And drug response. *N. Engl. J. Med.* **348(6):**529–537, 2003.
6. Kalow, W. and Gunn, D. R. Some statistical data on atypical cholinesterase of human serum. *Ann. Hum. Genet* **23:**239–250, 1959.
7. Evans, D. A., Manley, K. A., and McKusick, V. A. Genetic control of isoniazid metabolism in man. *Br. Med. J.* **5197:**485–491, 1960.
8. Kalow, W. *Pharmacogenetics: Heredity and The Response To Drugs*. W. B. Saunders, Philadelphia, PA, 1962.
9. Motulsky, A. G. Drug reactions enzymes, and biochemical genetics. *J. Am. Med. Assoc.* **165(7):**835–837, 1957.
10. Motuslky. A. G. From pharmacogenetics and ecogenetics to pharmacogenomics. *Med. Secoli* **14(3):**683–705, 2002.
11. Vogel, F. Modern problems of human genetics. *Ergib. Kinderh.* **12:**52–125, 1959.
12. Anzenbacher, P. and Anzenbacherova, E. Cytochromes P450 and metabolism of xenobiotics. *Cell. Mol. Life Sci.* **58(5–6):**737–747, 2001.
13. Wilkinson, G. R. Pharmacokinetics: The dynamics of drug absorption, distribution, and elimination, in Hardman, J. G. ed., *Goodman & Gilman's The Pharmacological Basis of Therapeutics,* McGraw-Hill, New York, 2004, 3–29.
14. Cupp, M. J. and Tracy, T. S. Cytochrome P450: new nomenclature and clinical implications. *Am. Fam. Physician* **57(1):**107–116, 1998.
15. Ingelman-Sundberg, M. Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, present and future. *Trends Pharmacol. Sci.* **25(4):**193–200, 2004.
16. Ma, M. K., Woo, M. H., and Mcleod, H. L. Genetic basis of drug metabolism. *Am. J. Health Syst. Pharm.* **59(21):**2061–2069, 2002.
17. Rushmore, T. H. and Kong, A. N. Pharmacogenomics, regulation and signaling pathways of phase I and II drug metabolizing enzymes. *Curr. Drug Metab.* **3(5):**481–490, 2002.
18. Brosen, K. and Gram, L. F. Clinical significance of the sparteine/debrisoquine Oxidation Polymorphism. *Eur. J. Clin. Pharmacol*. **36(6):**537–547, 1989.
19. Eichelbaum, M. and Gross, A. S. The genetic polymorphism of debrisoquine/sparteine metabolism—clinical aspects. *Pharmacol. Ther.* **46(3):**377–394, 1990.
20. Zanger, U. M., Fischer, J., Raimundo, S., et al. Comprehensive analysis of the genetic factors determining expression and function of hepatic CYP2D6. *Pharmacogenetics.* **11(7):**573–585, 2001.
21. Daly, A. K., Brockmoller, J., Broly, F., et al. Nomenclature for human CYP2D6 alleles. *Pharmacogenetics*. **6(3):**193–201, 1996.
22. Meyer, U. A. and Zanger, U. M. Molecular mechanisms of genetic polymorphisms of drug metabolism. *Annu. Rev. Pharmacol. Toxicol.* **37:**269–296, 1997.
23. Chen, S., Chou, W. H., Blouin, R. A., et al. The cytochrome P450 2D6 (CYP2D6) enzyme polymorphism: screening costs and influence on clinical outcomes in psychiatry. *Clin. Pharmacol. Ther.* **60(5):**522–534, 1996.
24. Griese, E. U., Zanger, U. M., Brudermanns, U., et al. Assessment of the predictive power of genotypes for the in-vivo catalytic function of CYP2D6 in a German population. *Pharmacogenetics.* **8(1):**15–26, 1998.
25. Dalen, P., Dahl, M. L., Ruiz, M. L., Nordin, J., and Bertilsson, L. 10-Hydroxylation of nortriptyline in white persons with 0, 1, 2, 3, and 13 functional CYP2D6 genes. *Clin. Pharmacol. Ther.* **63(4):**444–452, 1998.
26. Dalen, P., Dahl, M. L., Eichelbaum, M., Bertilsson, L., and Wilkinson, G. R. Disposition of debrisoquine in Caucasians with different CYP2D6-genotypes including those with multiple genes. *Pharmacogenetics* **9(6):**697–706, 1999.
27. Johansson, I., Lundqvist, E., Bertilsson, L., Dahl, M. L., Sjoqvist, F., and Ingelman-Sundberg, M. Inherited amplification of an active gene in the cytochrome P450 CYP2D locus as a cause of ultrarapid metabolism of debrisoquine. *Proc. Natl. Acad. Sci. USA* 1993; **90(24):**11, 825–11, 829, 1993.
28. Evans, W. E. and Mcleod, H. L. Pharmacogenomics—drug disposition, drug targets, and side effects. *N. Engl. J. Med.* **348(6):** 538–549, 2003.
29. Ingelman-Sundberg, M., Oscarson, M., and Mclellan, R. A. Polymorphic human cytochrome P450 enzymes: an opportunity for individualized drug treatment. *Trends. Pharmacol. Sci.* **20(8):** 342–349, 1999.
30. Chalon, S. A., Desager, J. P., Desante, K. A., et al. Effect of hepatic impairment on the pharmacokinetics of atomoxetine and its metabolites. *Clin. Pharmacol. Ther.* **73(3):**178–191,2003.
31. Sauer, J. M., Ponsler G. D., Mattiuz, E. L., et al. Disposition and metabolic fate of atomoxetine hydrochloride: the role of CYP2D6 in human disposition and metabolism. *Drug. Metab. Dispos.* **31(1):** 98–107, 2003.
32. Witcher, J. W., Long, A., Smith, B., et al. Atomoxetine pharmacokinetics in children and adolescents with attention deficit hyperactivity disorder. *J. Child. Adolesc. Psychopharmacol.* **13(1):**53–63, 2003.
33. Remy, C. N., Metabolism of thiopyrimidines and thiopurines. S-Methylation with S-adenosylmethionine transmethylase and catabolism in mammalian tissues. *J. Biol. Chem.* **238:**1078–1084, 1963.
34. Woodson, L. C. and Weinshilboum, R. M., Human kidney thiopurine methyltransferase. Purification and biochemical properties. *Biochem Pharmacol* **32(5):**819–826, 1983.
35. Ferrajoli, A. and O'brien, S. M. Treatment of chronic lymphocytic leukemia. *Semin. Oncol.* **31(2 Suppl. 4):**60–65, 2004.
36. Hia, C. P., Yip, W. C., Tai, B. C., and Quek, S. C. Immunosuppressive therapy in acute myocarditis: an 18 year systematic review. *Arch. Dis. Child.* **89(6):**580–584, 2004.
37. Shapiro, R. Low toxicity immunosuppressive protocols in renal transplantation. *Keio J. Med.* **53(1):**18–22, 2004.
38. Ardizzone, S. and Porro, G. B. Comparative tolerability of therapies for ulcerative Colitis. *Drug Safety* **25(8):**561–582, 2002.
39. de Jong, D. J., Derijks, L. J., Naber, A.H., Hooymans, P. M., and Mulder, C. J. Safety of thiopurines in the treatment of inflammatory bowel disease. *Scand. J. Gastroenterol.***(239)**69–72, 2003.
40. Coulthard, S. A. and Hall, A. G. Recent advances in the pharmacogenomics of thiopurine methyltransferase. *Pharmacogenom. J.* **1(4):**254–261, 2001.
41. Dubinsky, M. C., Lamothe, S., Yang, H. Y., et al. Pharmacogenomics and metabolite measurement for 6-mercaptopurine therapy in inflammatory bowel disease. *Gastroenterology* **118(4):**705–713, 2000.
42. Mcleod, H. L., Krynetski, E. Y., Relling, M. V., and Evans, W. E. Genetic polymorphism of thiopurine methyltransferase and its clinical relevance for childhood acute lymphoblastic leukemia. *Leukemia* **14(4):**567–572, 2000.
43. Relling, M. V., Hancock, M. L., Rivera, G. K., et al. Mercaptopurine therapy intolerance and heterozygosity at the thiopurine S-methyltransferase gene locus. *J. Natl. Cancer. Inst.* **91(23):** 2001–2008, 1999.
44. Lennard, L. Therapeutic drug monitoring of cytotoxic drugs. *Br. J. Clin. Pharmacol.* **52 (Suppl. 1):**75s–87s, 2001.
45. Marsh, S. and Mcleod, H. L. Cancer pharmacogenetics. *Br. J. Cancer* 2004; **90(1):**8–11, 2004.
46. Ambudkar, S. V., Dey, S., Hrycyna, C. A., Ramachandra, M., Pastan, I., and Gottesman, M. M. Biochemical, cellular, and pharmacological aspects of the multidrug transporter. *Annu. Rev. Pharmacol. Toxicol.* **39:**361–398, 1999.
47. Fromm, M. F., Kim, R. B., Stein, C. M., Wilkinson, G. R., and Roden, D. M. Inhibition of P-glycoprotein-mediated drug transport: a unifying mechanism to explain the interaction between digoxin and quinidine [see comments]. *Circulation* **99(4):**552–557, 1999.

48. Kim, R. B., Fromm, M. F., Wandel, C., et al.. The drug transporter P-glycoprotein limits oral absorption and brain entry of HIV-1 protease inhibitors. *J. Clin. Invest.* **101(2):**289–294 ,1998.
49. Borst, P., Evers, R., Kool, M., and Wijnholds, J. A family of drug transporters: the multidrug resistance-associated proteins. *J. Natl. Cancer. Inst*. **92(16):**1295–1302, 2000.
50. Brinkmann, U., Roots, I., and Eichelbaum, M. Pharmacogenetics of the human drug-transporter gene MDR1: impact of polymorphisms on pharmacotherapy. *Drug Discov. Today* **6(16):**835–839, 2001.
51. Choo, E. F., Leake, B., Wandel, C., et al. Pharmacological inhibition of P-glycoprotein transport enhances the distribution of HIV-1 protease inhibitors into brain and testes. *Drug. Metab. Dispos*. **28(6):**655–660, 2000.
52. Rao, V. V., Dahlheimer, J. L., Bardgett, M. E., et al. Choroid plexus epithelial expression of MDR1 P glycoprotein and multidrug resistance-associated protein contribute to the blood-cerebrospinal-fluid drug-permeability barrier. *Proc. Natl. Acad. Sci. USA* **96(7):**3900–3905, 1999.
53. Thiebaut, F., Tsuruo, T., Hamada, H., Gottesman, M. M., Pastan, I., and Willingham, M. C. Cellular localization of the multidrug-resistance gene product P-glycoprotein in normal human tissues. *Proc. Natl. Acad. Sci. USA* **84(21):**7735–7738, 1987.
54. Hoffmeyer, S., Burk, O., von Richter, O., et al. Functional polymorphisms of the human multidrug-resistance gene: multiple sequence variations and correlation of one allele with P-glycoprotein expression and activity in vivo. *Proc. Natl. Acad. Sci. USA* **97(7):**3473–3478, 2000.
55. Drescher, S., Schaeffeler, E., Hitzl, M., et al. MDR1 gene polymorphisms and disposition of the P-glycoprotein substrate fexofenadine. *Br. J. Clin. Pharmacol.* **53(5):**526–534, 2002.
56. Kim, R. B., Leake, B. F., Choo, E. F., et.al. Identification of functionally variant MDR1 alleles among European Americans and African Americans. *Clin. Pharmacol. Ther.* **70(2):**189–199, 2001.
57. Fellay, J., Marzolini, C., Meaden, E. R., et al. Response to antiretroviral treatment in HIV-1-infected individuals with allelic variants of the multidrug resistance transporter 1: a pharmacogenetics study. *Lancet* **359(9300):**30–36, 2002.
58. Ioannidis, J. P., Ntzani, E. E., Trikalinos, T. A., and Contopoulos-Ioannidis D.G. Replication validity of genetic association studies. *Nat. Genet.* **29(3):**306–309, 2001.
59. Rabe, K. F. State of the art in beta2-agonist therapy: a safety review of long-acting agents. *Int. J. Clin. Pract.* **57(8):**689–697, 2003.
60. Small, K. M., McGraw, D. W., and Liggett, S. B. Pharmacology and physiology of human adrenergic receptor polymorphisms. *Annu. Rev. Pharmacol. Toxicol.* **43:**381–411, 2003.
61. Dishy, V., Sofowora, G. G., Xie, H. G., et al. The effect of common polymorphisms of the beta2-adrenergic receptor on agonist-mediated vascular desensitization. *N. Engl. J. Med.* **345(14):**1030–1035, 2001.
62. Martinez, F. D., Graves, P. E., Baldini, M., Solomon, S., and Erickson, R. Association between genetic polymorphisms of the beta2-adrenoceptor and response to albuterol in children with and without a history of wheezing. *J. Clin. Invest*. **100(12):**3184–3188, 1997.
63. Tan, S., Hall, I. P., Dewar, J., Dow, E., and Lipworth, B. Association between beta 2–adrenoceptor polymorphism and susceptibility to bronchodilator desensitisation in moderately severe stable asthmatics. *Lancet* **350(9083):**995–999, 1997.
64. Tollman, P., Guy, P., Altshuler, J., Flanagan, A., and Steiner, M. A revolution in R&D; how genomics and genetics are transforming the biopharmaceutical industry. www.bcg.com, 2001
65. Nowlan, W. Human genetics. A rational view of insurance and genetic discrimination. *Science* **297(5579):**195–196, 2002.
66. Rothenberg, K. H. and Terry, S. F. Human genetics. Before it's too late–addressing fear of genetic information. *Science* **297(5579):** 196–197, 2002.
67. Wertz, D. C. Ethical, social and legal issues in pharmacogenomics. *Pharmacogenom. J.* **3(4):**194–196, 2003.
68. Ingelman-Sundberg, M. Pharmacogenetics: an opportunity for a safer and more efficient pharmacotherapy. *J. Intern. Med.* **250(3):**186–200, 2001.
69. Evans, W. E. and Relling, M. V. Moving towards individualized medicine with pharmacogenomics. *Nature* **429(6990):**464–468, 2004.
70. Slamon, D. J., Leyland-Jones, B., Shak, S., et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N. Engl. J. Med.* **344(11):**783–792, 2001.
71. Horton, J. Trastuzumab Use in breast cancer: clinical issues. *Cancer Control* **9(6):**499–507, 2002.
72. Ross, J. S., Linette, G. P., Stec, J., et al. Breast cancer biomarkers and molecular medicine. *Expert. Rev. Mol. Diagn.* **3(5):**573–585, 2003.

VII Applications of Molecular Diagnostics for Human Cancers

30 Molecular Pathogenesis of Human Cancer

William B. Coleman and Gregory J. Tsongalis

1. INTRODUCTION

Cancer represents a significant health problem worldwide. The successful curative treatment of almost every form of this disease depends on early diagnosis and, in the case of solid tumors, surgical resection with or without adjuvant therapy. Intensive research efforts during the last several decades have increased our understanding of carcinogenesis and have identified a genetic basis for the multistep process of cancer development *(1–3)*. In several human tumor systems, specific genetic alterations have been shown to correlate with well-defined histopathologic stages of tumor development and progression *(4,5)*. Although the significance of mutations to the etiological mechanisms of tumor development has been debated, a causal role for such genetic lesions is now commonly accepted for many human tumors. Thus, genetic lesions represent an integral part of the processes of neoplastic transformation, tumorigenesis, and tumor progression, and as such, they represent potentially valuable markers for cancer detection and staging *(6,7)*. Through the application of specific and sensitive molecular methodologies, the clinical laboratory of the future will be able to effectively screen populations at high risk for the development of cancer, potentially impacting the early detection and diagnosis of human cancers. In addition, development of new molecular diagnostic assays will expand the ability of clinicians to accurately stage tumor development, monitor progression of metastatic disease, and evaluate therapeutic outcome, facilitating the application of effective intervention strategies in the treatment of human tumors.

2. CANCER: A MULTISTEP GENETIC DISEASE

Cancer development is a multistep process through which cells acquire increasingly abnormal proliferative and invasive behaviors. Furthermore, cancer represents a unique form of genetic disease, characterized by the accumulation of multiple somatic mutations in a population of cells undergoing neoplastic transformation *(1–3,8,9)*. Several forms of molecular alteration have been described in human cancers, including gene amplifications, deletions, insertions, rearrangements, and point mutations *(9)*. In many cases, specific genetic lesions have been identified that are associated with the process of neoplastic transformation and/or tumor progression in a particular tissue or cell type *(3)*. Statistical analyses of age-specific mortality rates for different forms of human cancer predict that multiple mutations in specific target genes are required for the genesis and outgrowth of most clinically diagnosable tumors *(10)*. In accordance with this prediction, it has been suggested that tumors grow through a process of clonal expansion driven by mutation *(1,8,11,12)*. In this model, the first mutation leads to limited expansion of progeny of a single cell, and each subsequent mutation gives rise to a new clonal outgrowth with greater proliferative potential. The idea that carcinogenesis is a multistep process is supported by morphologic observations of the transitions between premalignant (benign) cell growth and malignant tumors. In colorectal cancer (and some other tumor systems), the transition from benign lesion to malignant neoplasm can be easily documented and occurs in discernible stages, including benign adenoma, carcinoma *in situ*, invasive carcinoma, and, eventually, local and distant metastasis *(13)*. Moreover, specific genetic alterations have been shown to correlate with each of these well- defined histopathologic stages of tumor development and progression *(5,14)*. However, it is important to recognize that it is the accumulation of multiple genetic alterations in affected cells, and not necessarily the order in which these changes accumulate, that determines tumor formation and progression. These observations suggest strongly that the molecular alterations observed in human cancers represent integral (necessary) components of the process of neoplastic transformation and tumor progression. From the clinical perspective, the process of accumulation of genetic alterations and mutations, over time, during neoplastic transformation, tumorigenesis, and tumor progression provides a window of opportunity for early detection, diagnosis, and intervention *(6,7,15)*. However, the selection of appropriate molecular markers will depend on the nature and temporal occurrence of the various genetic alterations that govern the establishment of a particular tumor type and the relationship between these genetic alterations and the histopathological features of the developing tumor.

3. MUTATIONS AND CANCER

Mutation is the ultimate source of variability for individual cells (and organisms) and is an essential component of the

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

process of natural selection *(16)*. Tumorigenesis can be viewed simply as a process of natural selection in which cells develop a growth advantage that allows them to proliferate and invade under conditions where other (normal) cells cannot, and the acquisition of this ability is driven by mutation. In other words, tumor development and progression represents a form of somatic evolution, at the ultimate expense of the host organism *(17)*. The idea that somatic mutation could significantly contribute to cancer development was suggested by Boveri early in the 20th century *(18)*. At about the same time, De Vries proposed that certain forms of radiation (Röntgen rays) might be mutagenic (17), suggesting that mutation rates could be influenced by exogenous factors. Evidence in support of the idea that multiple somatic mutations occur in and contribute to the stepwise process of neoplastic transformation and tumorigenesis has been provided by numerous investigators *(19–22)*. In early studies, the nature of the mutations and the contribution of these mutations to tumorigenesis were not at all clear. Nonetheless, the presence of multiple mutations in cancer cells could be observed in the form of karyotypic alterations and abnormal chromosome numbers in tumor cells *(23,24)*. More recent studies utilizing comparative genomic hybridization extended these observations by identifying both gross (cytogenetically detectable) and subtle chromosomal abnormalities in several types of human neoplasm *(25)*. Subsequently, numerous positive and negative mediators (proto-oncogenes and tumor suppressor genes) of cell growth and differentiation have been identified and characterized, defining the basic role for these critical genetic elements in neoplastic transformation and tumorigenesis *(2,3,26)*. Very recently, microarray-based gene expression studies have provided definitive evidence that cancer is ultimately a disease of abnormal gene expression patterns *(27–30)*. Somatic mutations occurring in developing cancers alter gene expression patterns, resulting in significant changes to cellular physiology, including unregulated (or abnormally regulated) cell proliferation and acquisition of invasive behaviors *(31,32)*. These investigations have shown that the gene expression signature of a specific cancer can be used in differential diagnosis, prognostication, and prediction of responses to therapy *(33,34)*.

The exact number of critical mutations required for neoplastic transformation of normal cells is not known. Investigations involving the statistical analysis of human tumor incidence and natural history in sporadic and inherited human tumors formed the basis for the two-hit model of cancer development *(35,36)*. In this model, genetic predisposition for a specific type of neoplasm is conferred on an individual who either inherits or otherwise acquires a germline mutation in one allele of a critical target (such as a tumor suppressor gene), constituting the first "hit." The second "hit" represents an acquired somatic mutation in the remaining normal allele of the critical gene. Accumulation of two hits alters (or eliminates) normal gene function in affected cells, which proliferate to form a tumor. Although the kinetics of tumor formation for some neoplasms are consistent with this model, it is now recognized that neoplastic transformation involves the mutational alteration or aberrant expression of multiple genes that function in cell proliferation or differentiation. Furthermore, epigenetic mechanisms can contribute to the multihit model of cancer induction through the silencing of critical genes *(37–39)*. In recent years, a reexamination of the number of critical mutations needed for cancer development has led to the suggestion that as many as six to eight mutations might be necessary for progression to an invasive tumor *(5,10)*. These analyses provide estimates of the numbers of mutations involving genes that control proliferation and differentiation of specific cell types that might be necessary for neoplastic transformation of that cell type. However, numerous lines of evidence support the suggestion that tumors are mutation-prone and/or accumulate large numbers of mutations *(12,40,41)*, and some investigators have estimated that tumor cells could contain thousands or tens of thousands of mutations *(42,43)*.

4. MOLECULAR ALTERATIONS IN CANCER

The molecular alterations occurring in cancer typically reflect mutations and can be categorized into two major groups: (1) chromosomal abnormalities and (2) nucleotide sequence abnormalities. There has been some debate in the literature as to which forms of mutation (chromosomal or nucleotide sequence) are more prevalent in cancer cells and/or constitute the foundations of the molecular mechanism of neoplastic transformation *(44)*. However, there is abundant evidence that representations of both of these major categories of genetic abnormalities exist in most tumor cells and that both significantly contribute to neoplastic transformation.

4.1. CHROMOSOMAL ABNORMALITIES Chromosomal alterations in cancer include the gain or loss of one or more chromosomes (aneuploidy), chromosomal rearrangements resulting from DNA strand breakage (translocations, inversions, and other rearrangements), and gain or loss of portions of chromosomes (amplification, large-scale deletion). The direct result of chromosomal translocation is the movement of some segment of DNA from its natural location into a new location within the genome, which can result in altered expression of the genes that are contained within the translocated region. If the chromosomal breakpoints utilized in a translocation are located within structural genes, then hybrid (chimeric) genes can be generated. The major consequence of chromosomal deletion (involving a whole chromosome or a large chromosomal region) is the loss of specific genes that are localized to the deleted chromosomal segment, resulting in changes in the copy number of the affected genes. Likewise, gain of chromosome number or amplification of chromosomal regions results in an increase in the copy numbers of genes found in these chromosomal locations.

4.2. NUCLEOTIDE SEQUENCE ABNORMALITIES Nucleotide sequence alterations in cancer include changes in individual genes involving single-nucleotide changes (missense and nonsense) and small insertions or deletions (some of which result in frameshift mutations). Single-nucleotide alterations that involve a change in the normal coding sequence of the gene (point mutations) can give rise to an alteration in the amino acid sequence of the encoded protein. Missense mutations alter the translation of the affected codon, whereas nonsense mutations alter codons that encode amino acids to produce stop codons. This results in premature termination of

translation and the synthesis of a truncated protein product. Small deletions and insertions are typically classified as frameshift mutations, because deletion or insertion of a single nucleotide (for instance) will alter the reading frame of the gene on the 3′ side of the affected site. This alteration can result in the synthesis of a protein that bears very little resemblance to the normal gene product or in the production of a abnormal/truncated protein resulting from the presence of a stop codon in the altered reading frame. In addition, deletion or insertion of one or more groups of three nucleotides will not alter the reading frame of the gene, but it will alter the resulting polypeptide product, which will exhibit either loss of specific amino acids or the presence of additional amino acids within its primary structure.

5. DNA DAMAGE AND MOLECULAR ALTERATIONS IN CANCER

Mutations in critical targets leading to neoplastic transformation can result from exogenous insults (carcinogens, radiation) or from endogenous mutagenic factors *(45)*. DNA damage can result from spontaneous alteration of the DNA molecule or from the interaction of numerous chemical and physical agents with the structural DNA molecule. Spontaneous lesions occur during normal cellular processes such as DNA replication, DNA repair, or gene rearrangement or through chemical alterations of the DNA molecule itself as a result of hydrolysis, oxidation, or methylation *(12,46)*. In most cases, DNA lesions create nucleotide mismatches that lead to point mutations. Nucleotide mismatches can result from the formation of apurinic or apyrimidinic sites after depurination or depyrimidation reactions, from nucleotide conversions involving deamination reactions, or, in rare instances, from the presence of a tautometric form of an individual nucleotide in replicating DNA. Deamination of nucleotide bases that contain exocyclic amino groups results in the conversion of cytosine to uracil, adenine to hypoxanthine, and guanine to xanthine *(47)*. However, the most common nucleotide deamination reaction involves methylated cytosines. The deamination of 5-methylcytosine, which results in the formation of thymine, accounts for a large percentage of spontaneous mutations in human disease *(48)*.

Interaction of DNA with physical agents, such as ionizing radiation (X-rays), can lead to single-strand or double-strand breaks through sission of phosphodiester bonds on one or both polynucleotide strands of the DNA molecule *(47)*. Ultraviolet (UV) light can produce cyclobutane pyrimidine dimers between adjacent pyrimidine bases on the same DNA strand. Less frequently, UV light produces non-cyclobutane-type pyrimidine dimers or 6-4 photoproducts between adjacent nucleotides in TC, CC, and TT pyrimidine dimers. Other minor forms of DNA damage caused by UV light include strand breaks and crosslinks *(47)*. Nucleotide base modifications can result from exposure of the DNA to various chemical agents, such as *N*-nitroso compounds and polycyclic aromatic hydrocarbons *(47)*. Among the numerous sites in the chemical structure of the nucleotides subject to modification by alkylating chemicals, the N^7 position of guanine and the N^3 position of adenine are the most frequently altered. DNA damage can also be caused by chemicals that intercalate the DNA molecule and/or crosslink the DNA strands *(47)*. Bifunctional alkylating agents can cause both intrastrand and interstrand crosslinks in the DNA molecule.

6. ARE CANCER CELLS PRONE TO MUTATION?

It is widely accepted that cancer cells accumulate numerous genetic abnormalities (consisting of chromosomal alterations and/or nucleotide sequence mutations) during the protracted interval between the initial carcinogenic insult and the outgrowth of a tumor. Although there is evidence that at least a portion of the genetic changes occurring in neoplasia are related to the underlying molecular mechanism of neoplastic transformation (*4,5,49)*, whether the myriad of genetic lesions found in cancer cells are the causes or consequences of neoplastic transformation continues to be the subject of debate *(50)*. In addition, some investigators have suggested that the intrinsic mutation rate in mammalian cells is insufficient to account for the many genetic changes observed in cancer cells, leading to the suggestion that an early (and possibly essential) step in neoplastic transformation is the development of a condition of hypermutability or genetic instability *(51,52)*. In the past, increased rates of mutation in preneoplastic or neoplastic cells would have been attributed to exposure of these cells to exogenous mutagenic agents. However, more recent analyses of the nature and frequency of mutations occurring in human neoplasms suggests that a significant proportion results from spontaneous mutational mechanisms *(53)*. This observation strengthens the suggestion that cancer cells might exhibit diminished capacities for surveillance and repair of DNA lesions, leading to increased rates of spontaneous mutation and/or increased susceptibility to mutation following exposure to some exogenous carcinogenic agent. An alternative argument suggests that increased rates of mutation are not necessary for the accumulation of large numbers of genetic lesions in cancer cells, but that selection of advantageous mutations is a more important feature of the process of tumorigenesis *(52,54)*.

6.1. SPONTANEOUS MUTATION RATES IN NORMAL CELLS The measured spontaneous mutation rate of mammalian cells depends on the exact experimental conditions employed and the nature of the cells and target sequence examined *(55)*. Somatic mutation rates have been determined for a variety of cultured cell types through examination of the spontaneous mutation frequency at one of several specific loci, such as the hypoxanthine–guanine phosphoribosyltransferase gene, the Na^+-K^+-ATPase gene, or the adenine phosphoribosyltransferase gene. Using the results from several of these studies *(56,57)*, the spontaneous mutation frequency at the hypoxanthine–guanine phosphoribosyltransferase locus can be estimated to be approx 2.7×10^{-10} to 1×10^{-9} mutations/nucleotide/cell generation in untransformed human cells. This is consistent with calculations made by others for this same locus, where the spontaneous mutation rate was estimated to be 1.4×10^{-10} mutations/nucleotide/cell generation *(58)*. The latter mutation rate is sufficient to yield approximately three mutations per cell over the life span of an individual, which may be too low to account for the number of mutations thought to be required for carcinogenesis. This observation led to the hypothesis that an early event in neoplastic transformation might involve an increase in the

spontaneous mutation rate in cells that are progressing through this multistep pathway *(58)*. Cells expressing the "mutator phenotype" accumulate mutations more rapidly than normal cells and would, therefore, be more likely to sustain mutations in critical genes required for enhanced growth and tumorigenesis *(59,60)*.

6.2. MUTATION RATES IN CANCER CELLS In many studies, the measured mutation rate in malignant cells is significantly higher than that of corresponding normal cells. In some cases, the elevated mutation rates were 100-fold higher than in untransformed cells *(61,62)*. Tumor cell lines that are deficient for DNA repair exhibit mutation rates that are 750-fold higher than that displayed by DNA repair-proficient tumor cell lines *(63)*. In addition, the rate of gene amplification in malignant cells is much higher than in normal cells *(64)*. However, other studies find no difference in the spontaneous mutation rate between normal and malignant cells *(56,57)*, or they suggest that selective pressures associated with clonal expansion of altered cells represent a much more important feature of carcinogenesis than a hypermutational phenotype *(54)*. Thus, some cancer cells might express a "mutator phenotype" and exhibit an enhanced mutation rate compared to normal cells *(41)*, whereas other cancers might exhibit multiple mutations in the absence of any appreciable increase in mutation frequency. These observations suggest the possibility that multiple molecular mechanisms are needed to reconcile the occurrence of multiple mutations in human cancers and the expression of a mutator phenotype with elevated mutational frequency in only a subset of these tumors.

7. GENOMIC INSTABILITY IN HUMAN CANCER

An appropriate definition of genomic instability is needed before a complete understanding of the interconnecting causes and consequences of genomic instability can be developed, and the contribution of this phenomenon to neoplastic transformation can be appreciated. The observation that most cancer cells contain discernible genetic abnormalities (chromosomal aberrations and/or DNA sequence abnormalities) suggests that all neoplastically transformed cells have sustained genetic damage and might have experienced some form of genomic instability during their development. Normal human cells demonstrate a remarkable degree of genomic integrity, which reflects the combined contributions of high-fidelity DNA replication processes, and the expression of multiple mechanisms that recognize and repair DNA damage. Nonetheless, it is recognized that rare spontaneous mutations can occur in cells that are proficient for both DNA replication and repair. The observation that neoplastic cells contain variable numbers of mutations reflecting specific forms of DNA damage and that tumors develop over widely variable periods of time (from initiation of the transformation process to the outgrowth of a clinically detectable tumor) suggests the possible involvement of different pathogenic mechanisms that might reflect multiple distinct mutagenic pathways to neoplastic transformation. Tumors are highly variable with respect to their growth characteristics; some tumors become clinically evident early in the human life span, whereas others present later in life. This discrepancy could reflect individual differences among tumors and tumor types with respect to the relative rapidity of their development and progression. Consistent with the proposal that tumors form through clonal expansion driven by mutation *(1,8,11)*, tumors displaying early-onset and rapid progression might accumulate a critical level of genetic damage more quickly than tumors with later onset and a more indolent course.

The forms of genetic damage typically displayed by cancer cells (involving chromosomal alterations and/or DNA sequence alterations) are not mutually exclusive. However, the evidence available suggests the involvement of different mutagenic mechanisms in the origins of these genetic abnormalities *(65–67)*. Nonetheless, it is likely that the same groups of target genes might be involved in tumorigenesis driven by the accumulation of either form of genetic damage. Inactivation of the p53 tumor suppressor gene (loss of function) can be accomplished through point mutation at numerous nucleotide sites *(26,68)* or through deletion of the locus on 17p *(69)*. Likewise, activation of proto-oncogene function can be accomplished by point mutation, as with the H-*ras* gene *(70)*, or by chromosomal translocation, as with the c-*myc* gene *(71)*.

Based on these observations, a unifying hypothesis is required to describe the possible mechanisms of genomic instability that can account for the disparate numbers of mutations (specific loci vs widespread mutation) and diverse nature of genetic damage (types of mutation) that characterize various human cancers. We have proposed that at least two broad categories of genomic instability might exist: (1) progressive (persistent) genomic instability and (2) episodic (transient) genomic instability *(72)*. Evidence supporting the existence of these forms of genetic instability has emerged from studies in bacteria *(73)*, and good examples of each of these forms of genomic instability have been identified in subsets of human neoplasms. Progressive or persistent instability defines an ongoing mutagenic process, with new mutations occurring in each cell generation, and is associated with cells that are compromised in their ability to safeguard the integrity of their genome. This form of genomic instability would be transmitted from cell generation to cell generation as a heritable trait *(73)*. For instance, tumor cells from patients with hereditary nonpolyposis colorectal cancer (HNPCC) exhibit progressive genomic instability, which is manifest as alterations in microsatellite sequences *(74–76)*. In contrast to progressive instability, episodic or transient instability describes sporadic genetic damage in cells that are otherwise proficient in the various pathways that govern genomic homeostasis. This form of instability is associated with tumors that contain specific mutations and/or chromosomal alterations in the absence of widespread damage to the genome. The transient mutator state might account for a large portion of adaptive mutations occurring in cells *(73)*. For instance, cells exposed to high levels of oxidative or nutritional stress might incur and accumulate adaptive mutations that enable the altered cells to thrive under highly selective conditions. These mutations might occur in cells in the absence of cell proliferation *(59,60)*, but they would facilitate clonal expansion of an altered clone in response to subsequent selection pressures *(52)*. Cells exposed to high levels of reactive oxygen species might accumulate mutations in this manner *(77,78)*. Numerous sporadic tumor types exemplify this form of

instability, including sporadic colorectal tumors of the tumor suppressor pathway *(79)*, or the microsatellite mutator pathway *(78,80)*. It can be envisioned that both chromosomal abnormalities and DNA sequence abnormalities could result from the expression of either of these forms of genomic instability during neoplastic transformation.

8. ABNORMAL DNA REPAIR CONTRIBUTES TO GENOMIC INSTABILITY AND CANCER PREDISPOSITION

The ability to repair damaged DNA is fundamental to all biological processes because damaged sites in the genome can be converted to permanent mutations during DNA replication. The susceptibility of a particular cell type to carcinogenesis is related to its relative abilities to metabolize genotoxic carcinogens and to repair damaged DNA *(81)*. Furthermore, susceptibility to genotoxic damage partially depends on the temporal relationship among DNA damage, DNA repair, and DNA replication *(82)*. It follows that there are aspects of several normal cellular processes that can indirectly contribute to mutation in normal cells, including (1) slow repair of damaged DNA in specific gene sequences and (2) timing of replication of specific genes *(83)*. DNA damage is repaired through one of several distinct pathways, including enzymatic reversal repair, nucleotide excision repair, and postreplication repair. An extensive review of each of these DNA repair pathways is beyond the scope of this chapter. Several excellent reviews are available for interested readers *(47,84)*.

Genetic alterations that affect normal DNA repair mechanisms necessarily lead to an accelerated accumulation of DNA damage and mutation in affected cells. Numerous genes have been identified that encode proteins involved with DNA repair and are required for the maintenance of the stability of the genome. Mutation of any of these genes might lead to genetic instability and a mutation-prone phenotype, contributing to the multiplicity of mutations observed in human tumors *(12,42)*. Evidence for this suggestion comes from studies of several rare genetic disorders identified in humans that involve dysfunctional DNA repair pathways. These disorders include xeroderma pigmentosum, Cockayne's syndrome, trichothiodystrophy, ataxia telangiectasia, Bloom's syndrome, and Fanconi's anemia. Of these disorders, xeroderma pigmentosum, ataxia telangiectasia, Bloom's syndrome, and Fanconi's anemia predispose affected individuals to the development of various malignancies when exposed to specific DNA-damaging agents. Patients with xeroderma pigmentosum display hypersensitivity to uv light and increased incidence of several types of skin cancer, including basal cell carcinoma, squamous cell carcinoma, and malignant melanoma *(85)*. Patients with ataxia telangiectasia exhibit hypersensitivity to ionizing radiation and chemical agents and are predisposed to the development of B-cell lymphoma and chronic lymphocytic leukemias *(86)*, and affected women demonstrate an increased risk of developing breast cancer *(87,88)*. Patients with Fanconi's anemia demonstrate sensitivity to DNA crosslinking agents and are predisposed to malignancies of the hematopoietic system, particularly acute myelogenous leukemia *(89,90)*. Patients with Bloom's syndrome demonstrate an increased incidence of several forms of cancer, including leukemia, skin cancer, and breast cancer *(91,92)*. These patients exhibit chromosomal instability manifested as abnormally high levels of sister chromatid exchange *(93)*.

The molecular basis of several of these genetic DNA repair deficiencies has been partially determined through genetic complementation analyses. Each complementation group identified represents a different genetic defect that eliminates a specific functional aspect of a DNA repair pathway. Seven complementation groups have been identified for xeroderma pigmentosum *(94)*, four complementation groups have been identified for ataxia telangiectasia *(95)*, and four complementation groups have been identified for Fanconi's anemia *(96)*. The molecular defect in Bloom's syndrome has been suggested to involve faulty regulation of DNA repair processes rather than faulty DNA repair enzymes *(97)*. The candidate Bloom's syndrome gene product is an enzyme with helicase activity *(98)*. Candidate genes for each of the xeroderma pigmentosum complementation groups have now been cloned. Each of these genes encode proteins involved with various aspects of DNA nucleotide excision repair, including proteins that function in the recognition of DNA damage and factors that couple the processes of transcription and repair *(99)*. A candidate ataxia telangiectasia susceptibility gene (termed *ATM*) has been identified, cloned, and characterized *(100)*. The *ATM* gene product is similar to several mammalian phosphotidylinositol kinases that are involved in mitogenic signal transduction, meiotic recombination, and cell cycle control *(101)*.

9. CHROMOSOMAL ABNORMALITIES IN CANCER

The majority of human cancers (including solid tumors, leukemias, and lymphomas) contain chromosomal abnormalities, consisting of either numerical changes (aneuploidy) and/or structural aberrations *(102,103)*. These two general types of chromosomal damage might reflect two distinct mechanisms of chromosomal instability *(9,104)*: (1) chromosome number instability and (2) chromosome structure instability. In some forms of cancer, chromosomal instabilities predominate over nucleotide sequence instabilities, suggesting that these mechanisms of genetic instability might not significantly overlap. Recent evidence suggests a genetic basis for chromosomal instability in cancer, involving mutational inactivation of certain types of gene in aneuploid tumors *(105)*.

Detailed karyotypic studies have been carried out on a large number of tumor types. Many of these studies have examined chromosomal alterations in leukemia and lymphoma *(102)*, partially reflecting the relative ease with which chromosomes can be prepared from these cancer cells. Traditional cytogenetic analyses of solid tumors are more difficult. Nonetheless, a substantial body of literature on the chromosomal aberrations of solid tumors has emerged *(25)*. Additional methods have also been applied to examination of chromosomal abnormalities in solid tumors *(103)*. A large number of studies have investigated allelic loss of heterozygosity in various human solid tumors using Southern analysis or polymerase chain reaction (PCR) *(106–110)*. Although these methods do not provide the same information as karyotypic analysis, the indication of large-scale deletions can be inferred from the loss of multiple markers on a specific chromosomal arm *(111)*. In addition,

flow cytometry is now widely employed for determination of tumor ploidy *(112)*, and fluorescence *in situ* hybridization (FISH) is used to examine specific chromosome numbers and alterations *(113,114)*. Spectral karyotyping combines FISH with a karyotypic analysis of chromosomes from tumor cells *(115)*, eliminating the requirement for chromosome banding in karyotype analysis. Derivative methods have been developed that facilitate chromosome banding analysis through FISH *(116)*. These powerful new techniques for chromosomal analysis have generated a revival of genetic investigations of cancer at the level of the chromosome. A detailed review of chromosomal alterations in human cancer is beyond the scope of this chapter. Several excellent reviews are available *(25,117,118)*.

9.1. INSTABILITY OF CHROMOSOME NUMBER

Numerical alterations of chromosomes can involve both loss of entire chromosomes or allelic losses, which might be accompanied by duplication of the opposite allele. This phenomenon results in the generation of a tumor with normal karyotype but an abnormal allelotype *(106)*. Several studies have produced evidence that suggest that tumors arising in various tissues share a common chromosome number instability and might lose a significant number (25–50%) of alleles during neoplastic transformation and tumorigenesis *(106,108,119,120)*. These large-scale genomic changes could be the result of some form of progressive chromosomal instability *(121,122)*. In support of this suggestion are studies showing that gains and losses of multiple chromosomes occur in aneuploid colorectal cancer cell lines 10-fold to 100-fold more frequently than in diploid cancer cell lines of the same histological subtype *(66,123)*. In other studies, the rate of loss of heterozygosity (LOH) at marker loci proximal to a selectable gene (adenine phosphoribosyl transferase) was increased 10-fold in colorectal cancer cell lines that exhibit proficiency of mismatch repair compared with cell lines that lack mismatch repair *(124,125)*. In addition to these results, numerous studies have combined to show that aneuploid cancers exhibit highly variable karyotypes *(102,126)*, suggesting that new chromosomal variations are produced in a progressive manner during tumor outgrowth and evolution.

The absence of chromosomal instability in diploid cancers and/or cancers that exhibit nucleotide sequence alterations argues against a nonspecific mechanism for chromosomal instability related to abnormal properties of neoplastic cells *(9)*. Further, the high rates of numerical chromosomal alterations in aneuploid cells do not simply reflect the ability of these cells to survive changes in chromosome number *(66)*. Likewise, tetraploid cells resulting from the fusion of diploid cancer cells retain a stable tetraploid chromosome number *(66)*, suggesting that the presence of a nondiploid chromosome number does not in and of itself precipitate progressive chromosomal instability. Rather, the evidence from the literature supports the existence of a specific form of genetic instability in cancer cells that results from dysfunction of normal chromosomal homeostasis producing numerical chromosomal abnormalities. Several possibilities have been investigated, including the involvement of (1) mutant p53 protein, (2) abnormal centrosomes, (3) abnormal mitotic spindle checkpoint function, or (4) abnormal DNA-damage checkpoint function *(9,122,127)*.

9.1.1. Inactivation of the p53 Tumor Suppressor Leads to Abnormalities of Chromosome Number

The p53 tumor suppressor protein has long been suggested to play significant roles in cell cycle progression and cell cycle checkpoint function in response to DNA damage *(82,128)*. The *p53* gene is commonly mutated in human cancers *(53)*, and these same cancers frequently exhibit abnormalities of chromosome number *(117)*. Thus, numerous studies have been performed in order to define the role of *p53* in the maintenance of chromosomal stability in normal cells and instability in neoplastic cells. In various forms of cancer, the presence of chromosomal abnormalities correlates with *p53* mutation *(129,130)*. Cells in culture often become aneuploid concurrent with mutation or inactivation of *p53 (131,132)*. These observations suggests that loss of p53 function leads to abnormal regulation of mitosis and segregation of chromosomes *(133)*. However, several other lines of evidence do not support a direct role for *p53* mutation in the genesis of this form of chromosomal instability. For example, development of aneuploidy occurs very early in the process of neoplastic transformation and tumorigenesis *(134)*, and *p53* mutation typically occurs later in the process *(135)*. In addition, some diploid tumor cell lines that exhibit a stable karyotype also contain mutant *p53 (136)*. These observations combine to suggest that loss of normal p53 function could contribute significantly to chromosomal instability in certain forms of cancer but does not represent the primary cause of this form of genomic instability.

9.1.2. Abnormal Centrosome Function Leads to Chromosomal Abnormalities

Aneuploid tumors demonstrate significant numbers of chromosomal imbalances, whereas such imbalances are rare in diploid or near-diploid tumors. The abnormalities of chromosome number observed in aneuploid tumors are consistent with a mechanism involving dysfunction of chromosome segregation during mitosis. Several lines of evidence support the idea that the integrity of the centrosome plays an integral role in the development of aneuploidy. Human tumors and tumor-derived cell lines have been characterized to contain abnormal numbers of centrosomes, abnormally sized and shaped centrosomes, and multipolar spindles in a number of human neoplasms, including tumors of the breast, lung, prostate, colon, pancreas, head and neck, bile duct, and brain *(137,138)*. In addition, the numbers of centrosomes was elevated in six of the seven (85%) aneuploid colorectal carcinoma cell lines evaluated, compared to diploid tumor cell lines, which displayed normal centrosome numbers *(123)*. Further, centrosome function was impaired in four of the five (80%) aneuploid colorectal cancer cell lines examined, whereas centrosome function was found to be intact in all diploid tumor cell lines *(123)*. These observations suggest that abnormal centrosome number and/or function are common among neoplastic cells that display aneuploidy and might represent an essential component of chromosome number instability in human cancers.

The mechanism leading to the formation of increased numbers of centrosomes in cancer cells remains undefined. However, abnormal centrosome number and function has been linked to the STK15 kinase in some cancers *(139,140)* and to a related kinase (PLK1) in others *(141)*. The STK15 gene was

found to be amplified in approx 12% of primary breast cancers and in cell lines derived from neuroblastoma and tumors of the breast, ovary, colon, prostate, and cervix *(140)*. Overexpression of STK15 (evidenced by immunostaining) was detected in 94% of invasive ductal carcinomas of the breast irrespective of histopathological subtype, suggesting that overexpression of this centrosome-associated kinase might be a common feature of breast cancers *(142)*. In addition, overexpression of STK15 was found in cell lines that lacked evidence of gene amplification, and ectopic expression of STK15 in near-diploid human breast epithelial cells produced centrosome abnormality accompanied by induction of aneuploidy *(140)*. An alternative mechanism suggests that mutational inactivation of *p53* or functional inactivation of *p53* through binding by mdm2 results in abnormal centrosome numbers and induction of chromosomal instability *(143,144)*. Furthermore, there is evidence that loss of *BRCA1* or *BRCA2* can lead to centrosome amplification and chromosome segregation dysfunction *(145,146)*. These studies combine to suggest that a number of different genes might contribute to centrosome function and homeostasis in normal cells and that inactivation or dysregulation of one or more of them can lead to abnormal centrosome number/function.

9.1.3. Aberrant Mitotic Spindle Checkpoint Function Leads to Aneuploidy The mitotic spindle checkpoint governs proper chromosome segregation by ensuring that chromatid separation does not occur prior to completion of alignment of all chromosomes along the mitotic spindle *(147)*. It follows that if the mitotic spindle checkpoint is defective, chromosome segregation during mitosis will occur asynchronously, potentially producing an unequal distribution of chromatids between the daughter cells *(147)*. Evidence supporting a role for aberrant mitotic spindle checkpoint function in the development of aneuploidy includes the observation that aneuploid cells respond inappropriately to agents that disrupt the spindle apparatus, such as colcemid. Normal cells respond to colcemid treatment by arresting in metaphase, whereas cells that display instability of chromosome numbers prematurely exit mitosis and initiate another round of DNA synthesis *(67)*. The hallmark of mitotic spindle checkpoint defect is the inability to inhibit entry into the S-phase when mitosis cannot be completed because of damage to the mitotic spindle *(148)*. Mutation or aberrant expression of genes that encode proteins involved in mitotic spindle checkpoint function can eliminate proper checkpoint function in tumor cells, contributing to development of aneuploidy. A number of these genes have now been identified *(149)*. Alterations in mitotic spindle checkpoint genes have been documented in several human cancers, including decreased expression of *hMAD2* in breast cancers *(150)*, and mutations in the *hBUB1* gene in colorectal cancers *(67,151)*. However, these mitotic spindle checkpoint genes are not implicated in all aneuploid cancers. Some aneuploid breast cancers lack mutations in *hBUB1* and exhibit normal mRNA expression levels *(152)*. Likewise, cancers of the respiratory tract, including head and neck cancers, small-cell lung carcinoma, and non-small-cell lung carcinoma, have not been shown to have significant numbers of mutations in *hBUB1* (*151,153,154)*, and sporadic tumors of the digestive tract rarely contain mutations of *hBUB1* or *hsMAD2 (155)*. The absence of mutations or significant alterations in expression of mitotic spindle checkpoint genes in aneuploid cells suggests that additional genes and/or mechanisms of checkpoint inactivation are operational in the majority of cancers that demonstrate chromosomal instability. Certain *p53* mutations have been described that are associated with gain-of-function and relaxed spindle checkpoint function in response to mitotic inhibitors, suggesting that both mutational inactivation of *p53* and dominant gain-of-function mutations in *p53* can contribute to genomic instability and aberrant chromosome segregation *(156)*. In addition, defective checkpoint function has been demonstrated in patients with ataxia telangiectasia who carry mutations of the *ATM* gene *(157)*. These studies combine to suggest that a variety of genes might function in normal control of the mitotic spindle checkpoint, and when mutated or aberrantly expressed, they could contribute to chromosomal instability through inactivation of the mitotic spindle checkpoint.

9.1.4. Abnormal DNA Damage Checkpoint Function Leads to Aneuploidy The DNA damage checkpoint represents the major cellular mechanism that guards against the replication of damaged DNA or entry of cells with DNA damage into mitosis. The types of DNA damage that elicit checkpoint activation include polymerase errors remaining after DNA replication and other forms of incompletely repaired DNA, damage resulting from exposure to exogenous genotoxins (ionizing radiation, chemical mutagens, and others), and damage related to endogenous genotoxic insult (such as reactive oxygen species). A number of genes have been implicated in the control of this checkpoint, including *p53 (128)*, *ATM (158)*, *BRCA1* and *BRCA2 (159)*, and some others *(9)*. Functional inactivation of one or more of these genes through genetic or epigenetic mechanisms could result in a genomic instability related to the loss of the DNA-damage checkpoint. Loss of this checkpoint might then lead to development of aneuploidy, directly resulting from abnormal segregation of damaged chromosomes *(9)*.

9.2. INSTABILITY OF CHROMOSOME STRUCTURE The majority of human cancers exhibit chromosomal abnormalities, including marker chromosomes with altered structure. It is generally accepted that many (if not the majority) of the alterations of chromosome structure occurring in cancer cells confer some selective advantage to the evolving tumor. Thus, accumulation of a critical number of chromosomal aberrations or development of specific chromosomal abnormalities might represent essential steps in the process of neoplastic transformation. Three general forms of chromosomal alteration are observed in cancer cells: (1) gene amplifications, (2) rearrangements and translocations, and (3) large-scale deletions.

9.2.1. Gene Amplification The amplification of specific chromosomal segments or genes have been documented in some cancers and in many cancer cell lines *(64,160)*, some of which involve cellular proto-oncogenes resulting in abnormal expression levels of the proto-oncogene products (*161)*. In general, gene amplification occurs late in tumorigenesis and is associated with tumor progression. It is the recognized mechanism through which many tumors acquire resistance to chemotherapeutic agents. Thus, gene amplifications can profoundly affect tumor behavior and can have prognostic significance for some

cancers. However, gene amplifications probably are not involved with early genetic alterations in preneoplastic lesions leading to neoplastic transformation. The mechanisms governing gene amplification have not been determined with any certainty. However, several studies suggest that gene amplification occurs at much higher rates in neoplastic cells than in normal cells *(64)*. A role for the *p53* tumor suppressor in gene amplification has been suggested by some investigators. Evidence supporting this suggestion includes the observation that gene amplification occurs more readily in cells following inactivation of *p53* function *(162,163)*. However, gene amplification can also occur in cells with normal *p53 (162)*. One possibility for the role of *p53* in this process is that amplification of a chromosomal segment in a normal cell might trigger apoptosis in response to perceived DNA damage *(164)*, whereas in the absence of normal *p53* function cells would not undergo apoptosis, but would continue to accumulate amplicons in subsequent rounds of replication *(9)*. Thus, this form of chromosomal instability might involve a mechanism (or a mechanistic component) that increases the ability of an affected cell to survive the genetic alteration.

9.2.2. Chromosomal Rearrangements and Translocations Chromosomal rearrangements can take on several different forms, the most common of which are chromosomal translocations. Patterns of chromosomal translocation in human cancer can be classified as complex or simple *(9)*. In some human cancers, no consistent pattern of chromosomal abnormality can be discerned (complex translocations). These tumors exhibit complex type translocations, which might appear to be random. Among individual tumors of one type, or individual cells of a single tumor, different chromosomal aberrations might be found. Very often, these rearrangements are accompanied by large-scale loss of chromosomal segments. Although it is possible that some of these chromosomal alterations are not essential to tumorigenesis, it is unlikely that any chromosomal alteration that does not confer a proliferative or adaptive advantage would be preserved in an evolving tumor. In some human cancers, specific chromosomal anomalies are consistently found in a high percentage of tumors (simple translocations). These recurrent chromosomal abnormalities might reflect molecular alterations that are essential and necessary to the molecular pathogenesis of the specific tumor type. The discovery of the Philadelphia chromosome [trans(9;22)(q34;q11)] in the cancer cells of patients with chronic myelogenous leukemia was the first report suggesting the involvement of nonrandom chromosomal changes in the molecular pathogenesis of the disease *(23)*. Subsequent studies suggest that the neoplastic cells of 80–90% of leukemia and lymphoma patients contain some sort of demonstrable karyotypic abnormality, and many of these are uniquely associated with morphologically or clinically defined subsets of these cancers *(102,126)*. Similar relationships between chromosomal alterations (and other genetic changes) and definable stages of tumor development and progression have been established for some human solid tumors, including colorectal carcinoma *(5,14)*, and proposed for others, including ovarian carcinoma *(165)* and pancreatic carcinoma *(166)*. The role of chromosomal translocation in cancer pathogenesis has been suggested to involve activation of proto-oncogenes by repositioning of the gene adjacent to a heterologous genetic control element. Evidence for this type of proto-oncogene activation includes studies of chromosome translocations in Burkitt's lymphoma *(167)*. In this cancer, the c-*myc* proto-oncogene is translocated from chromosome 8 to chromosome 14, proximal to the immunoglobulin enhancer sequences, resulting in abnormal constitutive expression of c-*myc (71)*.

9.2.3. Large-Scale Chromosomal Deletions Large-scale deletions of whole chromosomes or chromosomal arms have been documented in many cancers. These deletions can contribute to the abnormal allelotype of tumors and might accompany chromosomal rearrangements and/or translocations. In most cases, such deletions are thought to be related to the presence of a tumor suppressor locus on the affected chromosomal arm. Large-scale deletions affecting several chromosomes have been documented in sporadic colorectal carcinoma, including deletions of 5q, 17p, and 18q *(14)*. Each of these chromosomal arms contains a known tumor suppressor locus; the adenomatous polyposis coli (*APC*) gene at 5q *(168–171)*, the *p53* gene at 17p *(69,135)*, and the *DCC* (for "deleted in colorectal cancer") gene at 18q *(172)*.

10. MICROSATELLITE INSTABILITY IN HUMAN CANCER

Microsatellite instability is characterized by alterations to simple repeated sequences, including both expansions (insertions) and contractions (deletions), typically resulting in frameshift mutations. Microsatellites are repetitive sequences that consist of variable numbers of repeated units of one to four (or more) nucleotides. Such sequences are numerous and randomly distributed throughout the human genome *(173)*. Mutational alterations of numerous adenine mononucleotide repeat motifs (polyA tracts) distributed throughout the genome was the first characteristic used to define microsatellite instability in human tumors *(174,175)*. These early studies of sporadic colorectal cancers suggested that 12% of all tumors harbor these mutations, with as many as 1 x 10^5 mutated polyA tracts per tumor *(175)*. Subsequent studies demonstrated frequent microsatellite alterations in hereditary colorectal cancers *(74)* and sporadic tumors *(176)* when higher-order repeated units were examined.

10.1. DETERMINATION OF MICROSATELLITE INSTABILITY IN HUMAN TUMORS Tumors that express the microsatellite mutator phenotype contain numerous altered microsatellite sequences. However, not all microsatellite sequences are altered in all tumors expressing the mutator phenotype *(72,177)*. In fact, some studies have shown dramatic differences in susceptibility to mutation of individual microsatellite loci *(178–181)*. In addition, two distinct patterns of microsatellite alteration have been described in various human cancers that express the mutator phenotype, and specific microsatellite markers tend to be altered in a characteristic pattern *(179,182,183)*. Pattern 1 appears as a ladderlike banding pattern on electrophoresis, reflecting either expansion or contraction of the microsatellite repeat unit *(179,182)*. In contrast, pattern 2 shows a minor alteration of the microsatellite repeat unit, usually demonstrating a one-repeat unit expansion or contraction *(179,182)*. The pattern of alteration observed at a specific microsatellite locus might reflect the nature of the genomic

instability displayed by a tumor. Several factors might influence the probability of mutation at a specific microsatellite locus: (1) the type of repeated sequence (mononucleotide, dinucleotide, etc.), (2) the length of the microsatellite sequence (number of repeated units), (3) the location of the microsatellite sequence within the genome, and (4) the underlying molecular lesion. The results of several studies support these suggestions. Numerous polyA mononucleotide repeats have been evaluated in human cancers, and several of these have been shown to be useful in the determination of microsatellite instability *(178,184,185)*. However, some neoplasms that exhibit microsatellite instability demonstrate no polyA tract alterations and might only show alterations in higher-order repeats *(178)*. Thus, no single type of microsatellite will be diagnostic for microsatellite instability in all tumors. A direct relationship has been observed between the length of polyA tracts and their mutation frequency among these genetically unstable tumors *(186)*, consistent with the suggestion that the probability of sustaining a mutation in an individual microsatellite sequence is proportional to the length of its sequence *(187–189)*. Extensive comparison of the mutation of dinucleotide vs higher-order repeat units (trinucleotide or tetranucleotide) in human tumors supports the suggestion that larger alleles are more susceptible to mutation in genetically unstable tumors *(190)*. Studies with cancer cell lines that harbor mutations of known mismatch repair genes tend to demonstrate instability of specific classes of microsatellite. A cell line possessing a defect in *hPMS2* exhibits instability of trinucleotide repeats *(191)*, whereas other cell lines deficient for *hMSH3* or *hMSH6* demonstrate an inability to correct mismatches in dinucleotide (or higher-order) repeats *(192)*. Furthermore, cells lacking *hMSH6* (DLD-1 cells) demonstrate minimal levels of dinucleotide instability, whereas cell lines lacking *hMSH2* (LoVo cells) or *hMLH1* (HCT116 and SW48 cells) demonstrate profound dinucleotide instability *(193)*. In addition, the presence of mutations in specific mismatch repair genes can affect the extent of hypermutability at microsatellite sequences *(181)*. The microsatellite mutation rate in HCT116 cells is 10-fold greater than that of HEC-1-A cells *(181)*. The mismatch repair gene defects carried by these cell lines differ. HCT116 cells lack *hMLH1* and *hMSH3 (194–196)*, whereas HEC-1-A cells lack *hPMS2* and *hMSH6 (191,197)*. Thus, the severity of microsatellite instability might be related to the nature of mismatch repair gene mutations *(181)* as well as the number of these mutations (198). These observations suggest that the individual mismatch repair complexes might exhibit specificity for certain types of mismatchs and that the microsatellite mutator phenotype displayed by cancer cell lines might be directly related to the number and nature of mismatch repair gene mutations.

Despite the large number of published studies on microsatellite mutations in human cancer, there is no consensus on a definition of microsatellite instability. Numerous studies have described microsatellite instability that is limited to one locus in human tumors. However, it is clear that a more stringent definition of microsatellite instability is needed *(178,199,200)*. Generation of a uniform definition of microsatellite instability requires consideration of several important factors. These include (1) the number of microsatellites examined, (2) the identity of the microsatellite markers employed, and (3) the number of altered sequences per tumor. The reports of microsatellite instability in human cancers that are contained in the literature demonstrate a widely variable number of microsatellite markers examined. There is no convincing evidence that large numbers of markers (>20) are required to determine microsatellite instability in human tumors. Furthermore, it has been suggested that the absolute number of mutated microsatellite loci might not be as important as the overall frequency of unstable (altered) microsatellite markers *(178,179,200)*. Several lines of evidence suggest strongly that certain microsatellites are prone to mutation in specific types of cancer and, therefore, might be highly informative in these tumors. An example of this is the high frequency of mutation of the *BAT26* polyA tract in colorectal tumors *(184,185)*. Multiple studies have shown *BAT26* to be a highly informative marker of microsatellite instability in cancers of the colorectum and stomach as well as some others *(72)*. In fact, mononucleotide repeats (like *BAT26* and *BAT40*) are altered in >95% of mismatch repair-deficient cancers *(186)*. However, *BAT26* is not useful for the determination of microsatellite instability in some other tumors, most notably breast *(201,202)*. These observations combine to suggest that strategies for the determination of microsatellite instability in specific tumor types can be devised that combine a definition of instability (cutoff value for number of required alterations) and an appropriate battery of microsatellite markers (minimum number of specific microsatellite loci with high positive predictive value required for accurate prediction). Using this approach, recommendations for a uniform method for the determination of microsatellite instability in colorectal cancers have been developed by several laboratories *(178,203)* and through a consensus conference *(200)*. In one study, it was suggested that a combination of mononucleotide and higher-order repeats should be evaluated and that at least 40% of the loci should demonstrate instability for the tumor to be classified as unstable *(178)*. Furthermore, a group of 10 specific primer sets has been recommended that reliably detect microsatellite instability in colorectal tumors, including *APC*, *BAT25*, *BAT26*, *BAT40*, *D2S123*, *D10S197*, *D18S58*, *D18S69*, *mfd15*, and *MYCL1 (178)*. An alternative proposal recommends utilizing a reference panel consisting of mononucleotide and dinucleotide microsatellite markers (including *BAT25*, *BAT26*, *D2S123*, *D5S346*, and *D17S250*) for initial screening of colorectal tumors *(200,203,204)*. Using this panel of markers, the authors suggest that alteration of one marker indicates low-level microsatellite instability, a shift of two or more loci indicates high-level microsatellite instability, and genetically stable tumors exhibit no microsatellite alterations *(203)*. An additional number of microsatellite markers are recommended as alternative loci in the analysis of microsatellite instability in colorectal cancer *(200)*. When more than five loci are evaluated, alterations in at least 30–40% of the markers is required for an indication of high-level microsatellite instability *(200)*. A comparable strategy has been successfully employed in the assessment of microsatellite instability in ovarian cancers *(205)*. However, additional research is needed before similar diagnostic programs can be developed for the determination of microsatellite instability in other human tumors.

10.2. FREQUENCY OF MICROSATELLITE INSTABILITY IN HUMAN CANCER A large number of studies have now accumulated documenting the occurrence of microsatellite instability in human neoplasms. Many of these studies were recently reviewed *(72)*. Tumors from patients diagnosed with HNPCC frequently exhibit microsatellite instability (141/159 tumors, 89%) at 2 or more loci, whereas only 15% (887/5724 tumors) sporadic colorectal cancers demonstrate microsatellite instability at 2 or more loci *(72)*. However, sporadic colorectal cancers occurring in young patients (<35 yr of age) and in patients with multiple primary tumors exhibit microsatellite instability at higher frequency (64% and 71%, respectively) than sporadic colorectal tumors in general *(206)*. Sporadic gastric cancers exhibit microsatellite instability at 2 or more loci in 19% (276/1485) of tumors *(72)*, whereas gastric carcinomas from patients with multiple primary tumors or familial predisposition exhibit an elevated frequency of microsatellite instability (61% and 32%, respectively) compared to sporadic tumors *(206–208)*. Several other cancers exhibit microsatellite instability at multiple microsatellite loci in 15–35% of tumors examined. Sporadic breast cancers demonstrate microsatellite instability at 2 or more loci in 17% (64/372 tumors) of cases *(209–214)*, but this percentage varies widely from study to study. The combined results of 6 studies failed to detect microsatellite instability at even 1 locus among 522 tumors examined *(72)*, suggesting that the actual frequency of occurrence of microsatellite instability among breast cancers is very low. Microsatellite instability at 2 or more loci has been documented in 24% (168/713) of tumors of the endometrium, 13% (16/123) of ovarian cancers, 27% (25/92) of esophageal tumors, 28% (25/88) of liver tumors, 29% (78/272) of non-small-cell lung cancers, and 32% (79/247) of prostate cancers *(72)*. Hodgkin's disease and some forms of leukemia exhibit microsatellite instability in a high percentage of cases. However, additional studies will be needed to determine more precisely the prevalence of this genetic abnormality in these neoplasms, particularly among the various forms of leukemia. Microsatellite instability is rare (<10% tumors) among gliomas, neuroblastomas, and cancers of the testicles, thyroid, and uterine-cervix. Evidence for the involvement of microsatellite instability in some other tumors has been produced, although the number of tumors examined and the number of studies conducted are limited. Some tumor types exhibit extreme variability in the frequency of occurrence of microsatellite instability (among different studies or groups of tumors) or show no significant level of microsatellite instability.

10.3. MISMATCH REPAIR DEFECTS LEAD TO MICROSATELLITE INSTABILITY The molecular defects responsible for the microsatellite instability in human tumors involve the genes that encode proteins required for normal mismatch repair *(215)*. These include *hMSH2 (216,217)*, *hMSH3 (218)*, and *hMSH6/GTBP (219,220)*, which are human homologs of the bacterial *MutS* gene, and *hMLH1 (196,221)*, *hPMS1 (222)*, *hPMS2 (222)*, and *hMLH3 (223)*, which are human homologs of the bacterial *MutL* gene. One or more of these genes are mutated in the germline of the majority of individuals with hereditary nonpolyposis colorectal cancer *(196,216,217,221,222,224–227)*, and somatic mutations have been identified in sporadic colorectal tumors that display microsatellite instability *(179,228,229)*. Mismatch repair gene defects have also been identified in other forms of cancer that exhibit microsatellite instability *(65,72,230,231)*.

The proteins involved with mismatch repair operate in concert to recognize mispaired or unpaired nucleotides and facilitate their removal and repair *(230,232)*. This mechanism differs from nucleotide excision repair, which recognizes and repairs abnormal (adducts) nucleotides *(230)*. The observation that microsatellite mutations consist of expansion or contraction of the repeated sequence through insertion or deletion of variable numbers of repeat units has led to the suggestion that such mutations could arise through a slippage mechanism during replication of these simple repeat sequences *(233)*. Strand slippage of the primer at a repetitive sequence during replication generates a misaligned intermediate that is stabilized by correct base-pairing between discrete repeat units on the misaligned strand *(234)*. Such a misaligned intermediate is normally repaired through the proofreading function of the polymerase complex or by postreplication repair mechanisms *(235)*. If the intermediate is not repaired, subsequent rounds of replication will generate insertion or deletion mutations in the newly synthesized DNA strands. The relative location of the unpaired repeat sequence in the replication intermediate determines whether an insertion or deletion will result (contraction or expansion of the microsatellite). Evidence for the involvement of a misaligned replication intermediate has come from studies of mismatch repair in bacteria *(236)*.

10.4. MICROSATELLITE MUTATION RATES IN DNA REPAIR-PROFICIENT AND -DEFICIENT CELLS The spontaneous mutation rate of a dinucleotide microsatellite repeat sequence in normal human fibroblasts (NHF1 cells) has been estimated to be 12.7×10^{-8} mutations/cell/generation *(237)*, suggesting that dinucleotide repeat sequences are remarkably stable in normal human cells. This selectable reporter system has been employed to assess the mutation frequency of dinucleotide repeat sequences in tumor cell lines that exhibit mismatch repair defects. The mismatch repair-proficient HT1080 fibrosarcoma cell line exhibited a mutation rate of 9.8×10^{-6} mutations/cell/generation, whereas mismatch repair-deficient colorectal tumor cell lines H6 and LoVo gave mutations rates of 1.6×10^{-4} and 3.3×10^{-3} mutations/cell/generation, respectively *(238)*. The estimated mutation rates at the dinucleotide repeat in mismatch repair-deficient tumor cells were 16-fold (H6) and 337-fold (LoVo) higher than that of the mismatch repair-proficient HT1080 tumor cell line and 1260-fold and 25,984-fold higher than that of the normal NHF1 fibroblast cell line *(237)*. These results highlight the propensity for spontaneous mutation at microsatellite repeat sequences of tumor cells that are deficient for mismatch repair and support the notion of a mutator phenotype in cells that have sustained lesions in their mismatch repair systems.

10.5. SPECIFIC MISMATCH REPAIR GENES ARE RESPONSIBLE FOR SPECIFIC FORMS OF MICROSATELLITE INSTABILITY Genetic complementation studies have produced direct evidence for the involvement of specific chromosomal loci or specific genes in mismatch repair-deficient tumor cells that exhibit microsatellite instability. Transfer of human

chromosome 2, which contains the *hMSH2* and *hMSH6* genes, restores genetic stability and mismatch repair proficiency to *hMSH2*-mutant HEC59 endometrial tumor cells *(239)*. Furthermore, the chromosome 2 containing cells demonstrate microsatellite stability at a triplet repeat locus (*D7S1794*) and at a dinucleotide repeat (*D14S73*), whereas clones containing other transferred chromosomes (such as chromosome 17) continue to exhibit instability at these loci *(239)*. Furthermore, transfer of chromosome 2 restores genetic stability to HCT15 or DLD1 colon tumor cells, which carry mutations of both *hMSH6* and DNA polymerase δ *(220,240)*. This result suggests that the DNA polymerase δ defect is not the primary determinant of genetic instability in these cell lines *(239)*. However, other studies question whether *hMSH6* plays a major role in microsatellite instability *(241,242)*. Transfer of chromosome 3, which contains *hMLH1*, into HCT116 colon carcinoma cells that are homozygous for *hMLH1* mutation restores mismatch repair and stability to the *D5S107* dinucleotide microsatellite repeat *(243)*. In similar studies, transfer of human chromosome 5 (containing *hMSH3*) or human chromosome 2 into HHUA human endometrial tumor cells resulted in partial correction of the mismatch repair defect *(218)*. More recent studies have utilized single-gene transfer to correct mismatch repair deficiency. HEC-1-A endometrial carcinoma cells, which are defective for mismatch repair because of an *hPMS2* mutation *(191)*, show increased microsatellite stability, and reduced mutation rate at the hypoxanthine phosphoribosyltransferase locus, and cell extracts can perform strand-specific mismatch repair following transfection with a wild-type *hPMS2* gene *(244)*. Likewise, transfection of HCT15 colon tumor cells with *hMSH6* resulted in restoration of mismatch bind and repair capacities, increased stability of the *BAT26* polyA tract, and reduction in the mutation rate at the hypoxanthine phosphoribosyltransferase locus *(245)*.

10.6. EPIGENETIC SILENCING OF MISMATCH REPAIR GENES LEADS TO MICROSATELLITE INSTABILITY Mutational inactivation of mismatch repair genes have been documented in numerous human tumors that display microsatellite instability *(65,231)*. However, in many cases, the underlying molecular defect in mismatch repair cannot be identified in cells that display microsatellite instability, suggesting that additional mismatch repair genes exist or that alternative mechanisms for microsatellite mutation are operational in these tumors. Several studies have produced strong evidence that epigenetic regulation of mismatch repair gene expression might be responsible for loss of mismatch repair function in many tumors that display microsatellite instability. Initially, a strong correlation between general methylation status and mismatch repair proficiency in colorectal carcinoma cell lines was noted *(246,247)*. Cell lines that were deficient for mismatch repair and showed microsatellite instability demonstrated hypermethylation of endogenous and exogenous DNA sequences *(246,247)*. Subsequently, several laboratories have examined expression of *hMLH1*, methylation of the *hMLH1* promoter region, and microsatellite instability status among sporadic colorectal carcinomas *(248–250)*. Tumors exhibiting high-level microsatellite instability, no detectable expression of *hMLH1*, and no *hMLH1* point mutation, also showed hypermethylation of the *hMLH1* promoter region *(248–250)*. In cell lines that exhibit loss of *hMLH1* and hypermethylation of the *hMLH1* promoter, treatment with 5-aza-2′-deoxycytidine resulted in re-expression of *hMLH1* and restoration of mismatch repair capacity *(249)*. These results suggest that inactivation of *hMLH1* through hypermethylation of its promoter region might represent a principle mechanism of gene inactivation in sporadic colorectal carcinoma characterized by widespread microsatellite instability. Consistent with this suggestion, the *hMLH1* promoter has been shown to be hypermethylated in 122/167 (73%) colorectal carcinomas with microsatellite instability, but in only 20/138 (14%) of microsatellite stable colorectal cancers *(72)*. Likewise, *hMLH1* promoter hypermethylation is observed in 87/97 (90%) gastric tumors with microsatellite instability, but in only 14/265 (5%) of microsatellite stable tumors *(251,252)*. Although fewer tumors have been analyzed, this same relationship has been documented in endometrial carcinomas where 12/14 (86%) tumors with microsatellite instability show promoter hypermethylation vs 0/10 (0%) microsatellite stable tumors *(253)*. In contrast to the relationship observed in sporadic cancers with microsatellite instability, tumors from HNPCC patients that harbor mutations in mismatch repair genes do not show *hMLH1* promoter hypermethylation.

10.7. OXIDATIVE STRESS AND LOSS OF MISMATCH REPAIR FUNCTION Chronic inflammation is known to contribute to DNA damage related to excess levels of free radicals. Mismatch repair-proficient cells are protected from mutational alteration of microsatellite sequences after exposure to low levels of hydrogen peroxide *(80)*. However, mismatch repair-deficient cells demonstrate numerous microsatellite alterations in response to oxidative stress, and oxygen radical scavengers diminish the damaging effects of free radicals in these cell lines *(254)*. These observations suggest that DNA damage related to free-radical exposure can contribute to microsatellite instability. Many forms of cancer are closely associated with chronic inflammation, leading to the suggestion that oxidative stress might significantly contribute to DNA damage, elevating the risk for neoplastic transformation in affected tissues. Ulcerative colitis is an inflammatory bowel disease associated with increased risk for colorectal cancer *(255)*. Microsatellite instability is one of the features of ulcerative colitis and related preneoplastic lesions *(256)*, suggesting a role for mismatch repair defects in this condition. The accumulation of microsatellite mutations in ulcerative colitis could be related to a failure of the mismatch repair pathways to correct the excess damage resulting from elevated levels of free radicals *(257)*. Alternatively, mismatch repair function might be disabled through DNA methylation *(258)* or directly by oxidative stress *(259)*. The inactivation of mismatch repair function in response to oxidative stress is mediated by oxidative damage to mismatch repair complexes, possibly involving hMutSα, hMutSβ, and hMutLα *(259)*. This type of mechanism could account for microsatellite instability in chronically inflamed nonneoplastic tissues *(260)* as well as cancers associated with inflammatory processes *(261,262)*.

10.8. MOLECULAR TARGETS OF MICROSATELLITE INSTABILITY Numerous simple repeat sequences are found in the human genome. Some of these occur within the coding regions of structural genes. These genes might be targets for

mutation in cells that express the microsatellite mutator phenotype *(263,264)*. The *TGFβRII* gene contains two simple repeat sequences: (1) a 10-bp adenine mononucleotide tract at nucleotides 709–718 and (2) a 6-bp GT repeat at nucleotides 1931–1936 *(265)*. This gene represents the first recognized target for inactivation as a result of microsatellite mutations in human tumors and cell lines, and both of the simple repeat sequences contained within the gene have been shown to be subject to mutation *(266)*. Mutation of the $(GT)_3$ repeat region in one tumor by insertion of an additional GT repeat unit resulted in a frameshift that was predicted to significantly alter the carboxy-terminus of the receptor protein *(266)*. Additional mutations were documented in the $(A)_{10}$ repeat region of the *TGFβRII* gene (deletion of one or two bases) resulting in frameshifts that were predicted to give rise to truncated receptor proteins *(266)*. Inactivating *TGFβRII* mutations involving these simple repeat regions have now been identified in a significant number of human tumors that exhibit microsatellite instability, including sporadic and hereditary cancers of the colorectum as well as cancers of the stomach, endometrium, and acute lymphoblastic leukemia *(72)*. However, cancers of the esophagus *(267)* and gliomas *(268)* display no mutation in the microsatellite repeats of the *TGFβRII* gene.

Transforming growth factor (TGF)-β is a multifunctional protein that acts on a wide range of cell types through interaction with specific cell surface receptors *(269)*. Among its many functions, TGF-β has been shown to act as a dominant-negative growth regulator *(269)*. Thus, loss of TGF-β responsiveness could significantly contribute to uncontrolled cell proliferation and tumorigenesis in affected cells. The HCT116 colon carcinoma cell line expresses the microsatellite mutator phenotype and harbors a 1-bp deletion of the *TGFβRII* $(A)_{10}$ microsatellite repeat *(270)*. Reconstitution of a functional TGF-β type II receptor through transfection with a functional *TGFβRII* gene results in modulation of the neoplastic phenotype of these cells, including decreased clonogenicity in soft agar culture, and reduced/delayed tumorigenicity in athymic mice *(270)*. These results suggest that mutational inactivation of the *TGFβRII* gene might significantly contribute to the neoplastic phenotype of human colon tumors of the microsatellite mutator phenotype. Evidence suggesting a significant role for *TGFβRII* mutation in colorectal tumorigenesis was obtained through examination of the relationship among microsatellite instability, tumor grade, and *TGFβRII* mutation status in early and late colonic adenomas *(271)*. Mutations of *TGFβRII* were observed in high-grade dysplastic adenoma, but were absent from surrounding simple adenoma. These results suggest strongly that *TGFβRII* mutation represents a late event in colonic adenomas that express microsatellite instability and that these mutations correlate with progression from adenoma to carcinoma in the colon *(271)*.

In addition to *TGFβRII*, a number of other genes that function in various aspects of normal cellular homeostasis (growth control and DNA repair) exhibit frameshift mutations at microsatellite loci, including *APC*, *BAX*, *E2F-4*, *IGFIIR*, *hMSH3*, *hMSH6*, *TCF-4*, *BLM*, and others *(72,264,272,273)*. Mutation in these genes have been identified in a significant percentage (as high as 50–55%) of gastrointestinal cancers (HNPCC, sporadic colorectal, stomach) that exhibit microsatellite instability *(72)*. However, some other cancers that display microsatellite instability do not contain mutations of these particular genes *(274)*, suggesting that these genes might be preferential targets in tumors of the gastrointestinal tract. Other tumors with microsatellite instability might mutate different genes from those that have been identified to be susceptible to this form of genetic event *(264,273)*.

11. COLORECTAL CANCER: A GENETIC MODEL OF HUMAN TUMOR DEVELOPMENT AND PROGRESSION

A tumor progression model describing the genetically defined stages of colorectal tumorigenesis has been established *(4,5,28)* and has come to represent the paradigm for multistep tumor development. The genetic model for colorectal carcinogenesis *(4)* suggests that (1) tumors arise as a result of mutational activation of proto-oncogenes and mutational inactivation of tumor suppressor genes, (2) alteration of at least four or five genes is required for development of a malignant tumor, and (3) the total accumulation of genetic alterations and not their relative order of appearance determine the biological behavior of the developing tumor.

In colorectal tumor development, mutations of the K-*ras* proto-oncogene and the *APC* gene occur in early lesions, whereas alterations of the *p53* and *DCC* tumor suppressor genes represent late molecular events occurring in advance tumors. Approximately 50% of colorectal carcinomas and colorectal adenomas >1 cm in size harbor *ras* gene mutations *(29)*. The occurrence of *ras* gene mutations in early colorectal lesions suggests that alteration of this gene might represent an initiating event in the development of a large percentage of colorectal tumors. In contrast, 70–80% of colorectal carcinomas demonstrate allelic losses at 17p and 18q *(5)*, and significant numbers of tumors exhibited *p53* tumor suppressor gene mutations in addition to 17p chromosome deletions *(30)*. Epigenetic alterations have also been documented in colorectal carcinogenesis, including global hypomethylation of the genome of tumors early in progression *(31)*. Accumulation of genetic alterations is associated with histologic progression through the dysplasia–carcinoma continuum. Greater than 90% of early adenomas demonstrate one or zero alterations, 25% and 49% of intermediate and late adenomas (respectively) demonstrate at least two alterations, and 90% of carcinomas exhibit two or more alterations *(4)*.

12. CANCER AS A GENETIC DISEASE: LESSONS LEARNED FROM FAMILIAL CANCERS

The familial clustering of specific types of cancers among cancer-prone families has been recognized for many years, and the occurrence of such clustering supports the role for some common heritable factor among the cancer-susceptible members of cancer-prone families. Several familial cancer syndromes have been characterized through the study of rare families in which individual members are affected by the same type of cancer or a variety of cancer types with an early age of onset and a pattern consistent with a hereditary mechanism, resulting in the identification of candidate cancer susceptibility

genes. When these genes are altered in the germline of an individual, that individual carries a higher risk for the development of neoplastic disease. Whereas a number of families affected by such familial cancer syndromes is low, the underlying susceptibility genes that provide the genetic basis for cancer predisposition are often important in the development of the corresponding sporadic cancers.

12.1. RETINOBLASTOMA Retinoblastoma is a rare malignant tumor of the retina in infants and is the result of mutations in the *Rb1* tumor suppressor gene *(275)*, which encodes a multifunctional protein *(276)*. Approximately 40% of retinoblastoma cases are familial in origin, where one mutant allele of the *Rb1* gene is inherited and a second somatic mutation is later acquired in the other *Rb1* allele. Familial retinoblastoma is characterized by multiple and bilateral tumors. Nonhereditary or sporadic retinoblastoma, which accounts for 60% of all cases, results from somatic mutation of both *Rb1* alleles. In these cases, there is usually only a single unilateral tumor. The early age of onset and characteristics of bilaterality of familial retinoblastoma was explained by Knudson in 1971 *(36)*. Knudson's model proposed a "two-hit" molecular mechanism for tumor development. In this model, individuals displaying a genetic predisposition for developing retinoblastoma (inherited or otherwise) had acquired a germline mutation in one *Rb1* allele, which constituted the first "hit." The second "hit" represented an acquired somatic mutation inactivating the one remaining normal *Rb1* allele. The accumulation of two hits eliminated *Rb1* tumor suppressor gene function in affected cells that then proliferates to form retinoblastomas. Consistent with this model, individuals who inherit the first hit are more likely to develop a tumor, and tumor onset is more likely to occur early in life. Normal individuals are statistically less likely to accumulate the required number of hits and, thereby, are less likely to form a tumor. In addition, the late onset of sporadic tumors reflects the probabilistic nature of spontaneous mutations at a specific genetic locus.

12.2. LI–FRAUMENI SYNDROME In a detailed study of four families in which siblings or cousins were affected by rhabdomyosarcoma, Li and Fraumeni described what appeared to be an inherited predisposition to cancer development *(277)*. Within these extended families there was an unusually high incidence of early onset cancers, including breast cancer, soft tissue sarcomas, brain tumors, and leukemias *(277)*. Follow-up studies of members in these four families revealed new cancers in previously unaffected individuals, and a number of individuals developed second primary tumors *(278)*. In both cases, the patterns and types of cancers that developed were similar to that in the original study, supporting the idea of genetic susceptibility *(279)*. Segregation analysis among the individuals in these families demonstrated that the distribution of cancers was compatible with a rare autosomal gene with an approximate frequency of 0.00002 and a penetrance of 50% by age 30 yr and 90% by age 60 yr *(280)*. The cancer susceptibility gene in the Li–Fraumeni syndrome was subsequently suggested to be the *p53* tumor suppressor gene *(281)*. Germline mutations of the *p53* gene were found in affected individuals of Li–Fraumeni syndrome families in studies conducted by several laboratories *(281,282)*. To date, nearly 250 germline mutations in the *p53* gene have been identified in the Li–Fraumeni syndrome *(283,284)*. However, the *p53* tumor suppressor gene is not affected in all Li–Fraumeni syndrome families *(284)*. The genetic basis for the cancer-prone phenotype in Li–Fraumeni families without *p53* mutations is not known. However, it has been suggested that the p53 protein function among individuals in this particular grouping of families is compromised by virtue of protein–protein interactions with other cellular proteins, such as the *mdm2* gene product, which is the natural regulator of p53 function *(285)*.

12.3. HEREDITARY COLORECTAL CANCER Colorectal cancer is a fairly common disease worldwide, and particularly in populations from Western nations *(286)*, and a substantial fraction of colorectal cancers exhibit a genetic component. Genetic susceptibility to colorectal cancer can occur in association with familial adenomatous polyposis coli (FAP) or in the absence of FAP in a condition termed "hereditary nonpolyposis colorectal cancer" (HNPCC), also known as Lynch syndrome *(287)*. Genes associated with each of these conditions have been identified and characterized and have been determined to play important roles in sporadic colorectal tumors. The *APC* gene, which is located on chromosome 5q21 *(168–170)*, is responsible for FAP and has been implicated in the development of sporadic colorectal cancers *(171)*. Mutation of the *APC* gene over a large area of the coding region results in inactivation of the gene product. The mechanism of inactivation frequently involves the generation of frameshift or nonsense mutations, which lead to the production of truncated APC protein products. A correlation between the location of mutations in the *APC* coding sequence and the FAP phenotype has been observed *(288)*. Germline mutations between codons 1250 and 1464 are associated with profuse polyps, whereas mutations occurring in other regions are associated with sparse polyps. In addition, it has been suggested that certain truncated forms of the APC protein can interact with wild-type protein, giving rise to dominant negative inhibition of APC function *(289)*.

Hereditary nonpolyposis colorectal cancer is characterized by the occurrence of predominately right-sided colorectal carcinoma with an early age of onset and an increased risk for the development of certain extracolonic cancers, including cancers of the endometrium, stomach, urinary tract, and breast *(290)*. HNPCC is thought to account for <10% of colorectal cancers. However, precise determination of the contribution of HNPCC to colorectal cancer is difficult because of the heterogeneity in the clinical features of the syndrome. Genetic linkage studies demonstrated a close linkage of the disease to anonymous microsatellite markers on chromosome 2 without allelic deletion in two large HNPCC kindreds *(291)*. HNPCC has been determined to be related to defective DNA repair *(264,287)*. Tumors associated with HNPCC exhibit a form of genomic instability that represents a unique mechanism for a genomewide tendency for instability in short repeat sequences (microsatellites), which was originally termed the "replication error phenotype" *(74,175,176)* and is now simply referred to as microsatellite instability. This form of genetic instability and mutation is not limited to neoplasms occurring in HNPCC patients. Rather, phenotypically normal cells from these individuals also display characteristics of genomic instability manifest

as microsatellite mutations *(292)*. The molecular defect responsible for microsatellite instability in HNPCC involves the genes that encode proteins required for normal mismatch repair *(293)*.

12.4. MUIR–TORRE SYNDROME AND TURCOT'S SYNDROME Two additional familial cancer syndromes have been described that exhibit clinical features similar to that of HNPCC or FAP. The Muir–Torre syndrome is defined by the development of at least one sebaceous gland tumor and a minimum of one internal tumor, which is frequently colorectal carcinoma *(294)*. This syndrome shares several features with HNPCC syndromes Lynch I and Lynch II, including the occurrence of microsatellite instability in a subset of tumors *(295)*, suggesting the possible involvement of abnormal mismatch repair mechanisms in the genesis of these tumors. Turcot's syndrome is defined by the occurrence of a primary brain tumor and multiple colorectal adenomas *(296)*. The molecular basis for this syndrome has been suggested to involve mutation of the *APC* gene or mutation of a mismatch repair gene in tumors exhibiting microsatellite instability *(297)*.

12.5. HEREDITARY BREAST CANCER Hereditary predisposition to breast cancer is a causal factor in approx 5–10% of all cases. In the last decade, two genes, termed *BRCA1 (298)* and *BRCA2 (299)*, have been implicated in the etiology of familial breast cancer, and a third breast cancer susceptibility gene was suggested to exist *(300)*. Although additional breast cancer susceptibility or modifer genes might exist *(301)*, the direct evidence for *BRCA3* (at 13q21) is lacking *(302)*. The *BRCA1* gene has also been implicated in the hereditary predisposition to the development of ovarian cancer *(303,304)*. Approximately half of all families displaying a dominant predisposition to breast cancer and 80–90% of families in which multiple cases of breast and ovarian cancer occur have been suggested to harbor germline mutations in the *BRCA1* gene *(305,306)*. These mutations are highly penetrant, conferring a risk of about 90% of either breast or ovarian cancer by age 70. The actual frequencies of occurrence of breast and ovarian cancers vary among families. Epidemiological evidence suggests a model in which the majority of families are strongly predisposed to breast cancer but have only a moderately increased risk of ovarian cancer, with the remaining families displaying an equally strong predisposition to both breast and ovarian cancer *(303)*. In type 1 breast/ovarian cancer syndrome, the lifelong risk of developing breast cancer is very high (80–90%) and there is a moderate lifetime risk for developing ovarian cancer (30%). In type 2 breast/ovarian cancer syndrome, the lifetime risk for breast cancer development is somewhat lower (approx 70%), whereas the lifetime risk of ovarian cancer is increased to approx 85%.

Mutations in the *BRCA1* gene have been identified over a large area of the coding sequence of the gene, and a large percentage of the identified mutations result in truncated BRCA1 protein products with abnormal function or complete loss of function *(307,308)*. Gayther and colleagues evaluated germline *BRCA1* mutations with respect to breast and ovarian cancer frequencies among affected family members *(309)*. These investigators found that mutations in the 3′ one-third of the coding sequence of the *BRCA1* gene were associated with a lower proportion of ovarian cancers, suggesting a connection between the location of the mutation and the chances for developing either breast or ovarian cancer *(309)*. The *BRCA2* gene has also been shown to be linked to early onset familial breast cancers *(71)* and to be mutated in the germline of families with a high rate of breast cancers *(64)*. However, unlike *BRCA1*, the *BRCA2* gene has been shown to be mutated in rare cases of male breast cancer and is not associated with hereditary ovarian cancer *(64)*. All of the mutations identified in the *BRCA2* gene to date represent small deletions that result in fameshifts *(64)*.

12.6. FAMILIAL MELANOMA Approximately 8–12% of all cases of cutaneous melanoma are related to an inherited predisposition *(310)*. Individuals at high risk are often identified as having dysplastic nevus syndrome, a highly penetrant autosomal-dominant disorder characterized by predisposition to the development of malignant melanoma *(311)*. A variety of studies have contributed to the suggestion that genes on chromosomes 1, 6, 7, and 10 (and to a lesser extent, on 2, 3, and 11) are involved in the etiology of malignant melanoma *(310)*. Genetic linkage studies implicate at least two familial melanoma loci that are located at 1p36 and 9p21 *(312,313)*. The candidate 9p21 gene is the *MTS1* (or *CDKN2* or *CDK4I*) tumor suppressor gene that encodes the p16 cyclin-dependent kinase inhibitor that negatively regulates CDK4 *(314)*. This gene is mutated in 75% of all cell lines established from sporadic metastatic melanomas *(314,315)*. However, analysis of several kindreds has failed to consistently associate germline mutations of this gene with the occurrence of melanoma among these families *(315–317)*. Thus, other genes located at 9p21 might be involved in the development of this tumor. An alternative mechanism for inactivation of normal cell cycle progression control pathways in melanoma has been suggested by Dracopoli and colleagues, who have described a mutation in the *CDK4* gene that gives rise to a protein product that cannot be regulated by normal p16 protein *(318)*. These observations combine to suggest that abrogation of the p16–CDK4 regulatory pathway, through the mutational alteration of either gene *(319,320)*, represents a critical step in the pathogenesis of malignant melanoma. Viable candidate genes have not been identified on the other chromosomes implicated in melanoma progression, although there is evidence to suggest that progression of the tumor involves genes that regulate apoptosis and cell adhesion *(310,319,320)*.

12.7. OTHER GENETIC SYNDROMES ASSOCIATED WITH INCREASED CANCER RISK Several syndromes have been described that are characterized by congenital anomalies with varying phenotypic features and increased risks for specific forms of cancer. Von Recklinghausen neurofibromatosis or neurofibromatosis type 1 (NF1) is one of the most common autosomal dominant disorders of humans, occurring with a frequency of 1 in 2500 to 1 in 5000 individuals *(321)*. NF1 is characterized by benign skin lesions (cafe au lait spots), Lisch nodules of the iris, and multiple benign cutaneous and soft tissue neurofibromata *(321)*. Individuals with NF1 are at high risk for the development of tumors of the central nervous system, malignant peripheral nerve sheath tumors, rhabdomyosarcoma, and leukemias. The genetic basis for NF1 has been determined to be a tumor suppressor gene located on 17q11.2 *(321,322)*. The *NF1* tumor suppressor gene *(323–325)* encodes a protein

with similarity to cytoplasmic GTPase-activating proteins that function in signal transduction *(326)*. Analysis of a large number of NF1 patients for mutations in the *NF1* gene demonstrate that inactivation occurs through multiple forms of molecular alteration, including point mutation and deletion *(327,328)*. Multiple endocrine neoplasia type 2 (MEN2) is characterized by medullary thyroid carcinoma, bilateral pheochromocytoma, and hyperparathyroidism *(329)*. MEN2 and familial medullary thyroid carcinoma are caused by germline mutations of the *RET* proto-oncogene *(330)*. Wilms' tumor or nephroblastoma is a pediatric kidney neoplasm that can develop as a sporadic tumor or in the setting of genetic predisposition. Genetically predisposed individuals include those affected by the WAGR syndrome (characterized by Wilms' tumor, aniridia, genitourinary malformations, and mental retardation), Denys–Drash syndrome (characterized by intersexual disorders, nephropathy, and Wilms' tumor), or Beckwith–Weidemann syndrome (characterized by macroglossia, organomegaly, hemihypertrophy, neonatal hypoglycemia, and various embryonal tumors) *(331)*. Initially, this pediatric tumor was thought to be related to the deletion of the *WT1* tumor suppressor gene *(332,333)*, located on human chromosome 11p13. It has since been recognized that Wilms' tumor constitutes a contiguous genetic syndrome whereby several genes, responsible for the various phenotypes of WAGR, Denys–Drash, and Beckwith–Weidemann syndromes, located in the same region of chromosome 11 are mutated or deleted *(331)*. A recent review showed that 40% of patients with Wilms' tumor harbor germline mutations in the *WT1* gene and that these patients have an earlier age of disease onset compared to patients lacking germline mutations *(334)*. In addition, specific *WT1* mutations have now been associated with specific features of the clinical syndromes that involve Wilms' tumor *(334)*.

13. MOLECULAR ALTERATIONS IN HUMAN TUMORIGENESIS: IMPLICATIONS FOR THE EARLY DETECTION AND DIAGNOSIS OF HUMAN CANCER

Exploitation of genetic alterations for the early detection and diagnosis of cancer prior to clinical manifestation should increase the survival of patients by allowing intervention at a time when lesions are either localized and most amenable to surgical resection, or not yet morphologically diagnosable *(335)*. Several technical considerations must be given to the appropriateness of a genetic test for early detection of cancers or preneoplastic lesions. These include (1) the genetic target or targets for a specific cancer type, (2) the appropriate source of patient DNA for testing procedures, (3) the nature of the mutational spectra for specific genetic targets observed in different forms of cancer, and (4) the sensitivity of the molecular methods applied.

The appropriate molecular targets for diagnostic procedures depends on the specific cancer subtype and the known genetic alterations associated with its initiation and progression into a clinical cancer. A significant component of this consideration is the frequency with which a specific target is mutated or altered during the process of neoplastic transformation. Following the identification of an appropriate genetic target, a source of patient DNA must be obtained for use in the testing procedure. Desirable sources for preparation of DNA include those that can be obtained with minimal invasion of the patient (such a blood, sputum, or stool). Once the DNA sample enters the molecular biology laboratory, it is necessary to consider the expected mutational spectrum with respect to where mutations are likely to be found within the overall structure of the gene. Genes that are typically mutated at a few hot-spot codons are preferable targets to genes that can be activated or inactivated through mutations affecting single nucleotides over a wide area of the gene sequence. The economic feasibility of the diagnostic procedure will depend greatly on this consideration. Assays that detect rare mutations (mutation frequency $>1 \times 10^{-6}$) might be too sensitive to be generally useful in clinical diagnosis *(7)*. Such tests have the potential to detect mutations that are fixed in a small number of cells prior to clonal expansion and clinical progression. Thus, a significant rate of false positives could occur, resulting in the inappropriate clinical screening or intervention in affected patients. The ultimate goal for the development of effective molecular screening techniques is the identification of molecular alterations prior to the onset of clinical symptoms or morphologic manifestations, such that patients identified in this manner could then be evaluated at regular intervals for the appearance of lesions that are curable by surgical resection. In addition, effective screening strategies could facilitate the identification of candidates for chemopreventative strategies.

13.1. EARLY DETECTION OF COLORECTAL CANCERS

Traditional screening methods for colorectal cancer through the use of stool guaiac tests remain controversial, with many false-positive and false-negative results *(336)*. With the establishment of the early genetic alterations in the development of colorectal cancer, it became evident that these changes could serve as markers for the early detection of developing neoplasms. This led to the discovery that *ras* gene mutations could be identified in DNA isolated from stool samples obtained from patients with colorectal cancer using a PCR-based method *(337)*. Subsequently, Tobi et al. *(338)* developed an enriched PCR method for detection of K-*ras* gene mutations in high-risk patients prior to the development of colorectal cancer. In their study, mutations in codon 12 of K-*ras* were detected in approx 40% of high-risk patients *(338)*. Subsequently, a number of investigators have demonstrated the utility of mutation analysis using stool-derived DNA samples to detect occult colorectal carcinoma. These studies have focused on K-*ras*, *p53*, and *APC* gene mutations *(339–341)*. More recently, Jen and colleagues examined a combination of mutations involving *p53*, *BAT26*, and K-*ras* in stool-derived DNA samples from colorectal cancer patients *(342)*. Together, these three genetic markers detected 71% of colorectal cancers (among all patients examined) and 92% of tumors in patients with one or more genetic alterations *(342)*. Application of methods like these might facilitate the identification of patients with developing colorectal neoplasms prior to the emergence of a tumor that can be identified using conventional methods (such as endoscopy, barium enema, or stool guaiac). Given that most colon cancers progress through the adenoma–carcinoma sequence, early detection will facilitate surgical intervention when adenomas are small and

localized, without high-grade dysplasia or carcinoma *in situ*, potentially impacting long-term patient survival.

Familial adenomatous polyposis is an inherited disease characterized by multiple colorectal tumors. Traditional diagnosis relies on detection of numerous colorectal polyps during young adulthood. Identification of the FAP locus at 5q21 and subsequent cloning of the *APC* gene *(168,170,171)* have provided the opportunity for genetic testing for this disease. However, the *APC* gene represents a difficult target for use in molecular diagnostics, because it can be inactivated through mutational alteration at a number of sites within its coding sequence, which encompasses more than 8500 bp of DNA *(168,170)*. Based on the observation that mutations in the *APC* gene frequently give rise to truncated gene products, Kinzler and colleagues developed a strategy for identification of mutations based upon examination of APC proteins synthesized in vitro and allele-specific expression of endogenous *APC* transcripts *(343)*. The protein assay revealed altered (truncated) APC protein products in 82% of patients evaluated. In patients who did not demonstrate aberrantly sized APC protein, the allele-specific expression assay showed reduced expression from one allele, suggesting mutational alteration of the affected allele. When applied in combination, these procedures were able to identify germline mutations in the *APC* gene in 87% of patients evaluated *(343)*. Recently, methods for genotyping patients using mass spectroscopy have been described *(344)*. Application of this method to *APC* genotyping facilitates high-throughput screening of patients and molecular diagnosis of FAP *(345)*. Application of molecular methods should enable the diagnosis of FAP before establishment of benign colorectal polyps and improve the management of these patients with respect to preventive measures prior to the development of colorectal tumors. The application of potential pharmacologic treatments for polyposis might be more effective if initiated prior to the appearance of polyps *(343)*.

13.2. EARLY DETECTION OF LUNG CANCER Sidransky and colleagues have examined the occurrence of K-*ras* and *p53* gene mutations in primary lung carcinomas and corresponding sputum samples obtained from the same patients prior to clinical diagnosis of their tumors *(346)*. In their study, 67% of tumors evaluated contained either a K-*ras* or *p53* gene mutation. Using a PCR-based assay, the same mutations were identified in sputum samples corresponding to 80% of cases. Each of these sputum samples was cytologically negative for diagnosis of carcinoma. This technique is extremely sensitive; a mutation was detected in a sputum sample of one patient more than 1 yr prior to the clinical diagnosis of lung carcinoma *(346)*. This method combines PCR amplification of the target gene with plaque screening of bacteriophage clones of the amplified DNA product using mutant or wild-type sequence-specific oligonucleotide probes. Subsequently, mutant gene sequences are confirmed through DNA sequence analysis. No K-*ras* or *p53* gene mutations were detected in five control lung tumors and these tumors were found to be negative by PCR cloning and plaque screening *(346)*, suggesting that this molecular assay is extremely specific. In addition, 75% of patients who initially tested positive for mutation were found to test negative in sputum samples obtained following complete surgical resection of their tumor. The authors suggest that sensitive molecular analyses can detect potentially diagnostic mutations in cytologically negative sputum samples prior to the clinical detection of lung cancer by radiological methods, presenting the possibility that chemopreventative strategies could be employed in the clinical intervention *(346)*.

Similar molecular approaches have been applied to plasma-derived DNA samples from patients with non-small-cell lung cancer, examining molecular alterations in *p53* and *FHIT* and microsatellite sequences from chromosome 3 *(347)*. Mutations in *p53* were detected in the plasma of 73% of patients harboring a p53-mutant tumors, alterations at *FHIT* and 3p loci were identified in plasma from 57% of patients with the same alterations in their tumor *(347)*. Alteration of at least one of these genetic targets was found in plasma samples from 52% of all patients and in 69% of patients with a stage 1 tumor *(347)*. In a similar study, K-*ras* mutations were investigated in plasma-derived DNA samples from non-small-cell lung cancer patients *(348)*. Mutations in K-*ras* were detected in plasma of 20% of patients, with identical mutations in the corresponding tumors. Evaluation of plasma DNAs after treatment revealed the loss of detectable K-*ras* mutations in patients that showed a complete response to therapy, whereas mutant K-*ras* continued to be detected in patients that did not respond well to therapy *(348)*. These results suggest that plasma DNA might be useful in non-invasive strategies for early detection of lung cancers and/or monitoring of response to therapy.

13.3. EARLY DETECTION OF PANCREATIC CANCER The K-*ras* proto-oncogene is activated by point mutation in 72–100% of primary pancreatic adenocarcinomas *(349)*. Thus, evaluation of mutations in the K-*ras* gene might facilitate the early diagnosis of pancreatic tumors. Bernaudin and colleagues developed a rapid screening method for K-ras mutations in fine-needle aspirates and evaluated the potential of K-*ras* mutations in the diagnosis of pancreatic adenocarcinoma *(350)*. In their method, regions of the K-*ras* gene were PCR-amplified using sequence-specific oligonucleotide primers, and a unique restriction site was introduced into the amplified products of a codon 12 mutant K-*ras* gene by virtue of the sequence-specific oligonucleotide primers utilized. Subsequently, products of mutant and wild-type K-*ras* genes were distinguished by restriction digestion and RFLP analysis. Using this methodology, a mutation present in only 1% of the cells within a given specimen could be detected. Mutations of the K-*ras* gene were detected in 11/12 (92%) pancreatic adenocarcinomas using PCR/RFLP, suggesting that evaluation of K-*ras* mutations could facilitate the diagnosis of pancreatic tumors *(350)*. In a similar study, Pradayrol and colleagues utilized PCR/RFLP to evaluate K-*ras* mutations in DNA obtained from samples of pancreatic juice collected during endoscopic retrograde pancreatography of patients prior to clinical diagnosis *(351)*. The results of this study suggested that evaluation of K-*ras* mutations in samples of pancreatic juice is useful in the differential diagnosis of neoplastic and non-neoplastic pancreatic disease *(351)*. Further, these investigators demonstrated that K-*ras* mutations can be detected in the pancreatic juice many months (18–40 mo) prior to the clinical manifestation of pancreatic cancer *(351)*. Other studies have also demonstrated the potential for

early detection of pancreatic cancers by evaluation of K-*ras* gene mutations in gastric aspirates and stool specimens *(352,353)*.

14. MOLECULAR ALTERATIONS IN HUMAN TUMORIGENESIS: IMPLICATIONS FOR THE PREDICTION OF CLINICAL OUTCOME

14.1. MOLECULAR PROGNOSTICATION OF COLORECTAL CARCINOMA The genetic alterations associated with colorectal carcinoma have been established *(4,5)*. Several groups have examined the value of these progression-associated genetic alterations in the prediction of clinical outcome. Smith and colleagues analyzed *p53* tumor suppressor gene mutations in colorectal cancer patients and correlated the presence of specific mutations with clinical aggressiveness, patient response to postoperative therapy, and patient survival *(354)*. In this study, patients found to possess a point mutation in the *p53* gene showed a significantly poorer prognosis than those without point mutation. In a similar study, Hamelin and colleagues observed a strong correlation between the presence of *p53* gene mutations and short survival among colorectal cancer patients *(103)*. Among patients with a mutated *p53* gene, the occurrence of the point mutation within the highly conserved regions of the gene demonstrated a significant association with lymph node dissemination and an increased risk for the development of distant metastasis, and patients possessing a mutation in codon 175 of the *p53* gene demonstrated the poorest prognosis of all patients *(354)*. In addition, the response of patients to palliative therapy was determined to be markedly dependent on the *p53* mutation status of the primary tumor. None of the patients with *p53* gene mutation survived to the 40-mo follow-up among those receiving palliative postoperative therapy, whereas only one patient without *p53* gene mutation died during the same interval *(354)*. These authors suggest that knowledge of the *p53* mutation status and the nature of specific lesions is required to predict accurately the clinical course of the disease, and the response of patients to postoperative therapeutic options *(354)*.

Hamilton and colleagues have examined the prognostic value of allelic losses at 18q in colorectal cancers *(355)*. Alterations at this chromosomal site generally represent late events preceding metastasis *(4,5)*. Thus, the authors hypothesized that allelic loss at 18q might be indicative of rapidly progressing cancers irrespective of the clinico-pathological diagnosis *(355)*. The experimental approach involved PCR amplification of polymorphic markers along chromosome 18q and PCR-mediated examination of chromosome-specific polymorphic microsatellite markers along chromosome 18q. Patients with Stage II disease and no 18q loss exhibited a significantly better prognosis than patients with Stage II disease and 18q loss (5-yr survival rates of 93% and 54%, respectively). In patients with Stage III disease, those with 18q loss had a 5-yr survival of 38% vs 52% for those without 18q loss. The authors observed that the prognosis of Stage II patients with 18q loss was comparable to that of Stage III patients without 18q loss *(355)*. Patients with Stage III disease are frequently treated more aggressively than Stage II patients, with routine clinical intervention involving both surgery and adjuvant therapy *(13)*. Thus, evaluation of Stage II patients for 18q allelic loss might identify a subset of patients who would benefit from aggressive clinical intervention *(355)*.

14.2. MOLECULAR PROGNOSTICATION OF BREAST CANCER Mutations of the *p53* tumor suppressor gene are variably but frequently observed in breast carcinoma (15–40%), suggesting that mutation of this gene is important in the genesis of this neoplasm *(356)*. The observation that patients with Li–Fraumeni syndrome exhibit an increased risk for development of breast tumors *(284)* supports this suggestion. Several studies have examined the prognostic significance of *p53* mutations in breast tumors. Kovach and colleagues determined that breast cancer patients with mutated *p53* exhibit a shorter time to tumor reoccurrence and a shorter overall survival than patients that do not possess a mutated *p53* gene *(357)*. Further, these investigators demonstrated that both missense-type mutations (point mutation and in-frame microdeletions) and null mutations (hemizygous nonsense and frameshift mutations) were associated with poor prognosis *(357)*. Thorlacius et al. *(358)* found that *p53* mutations occurred in approx 17% of breast carcinomas examined, but that patients with mutated *p53* exhibited significantly greater mortality rates than patients lacking *p53* mutations. Sommer and colleagues *(359)* found that detection of *p53* gene mutations was associated with a high probability of tumor recurrence and death in affected patients. These investigators suggest that evaluation of *p53* gene mutations in breast carcinomas could serve to identify patients who lack conventional indicators of poor prognosis, but who are at high risk for recurrence and early death *(359)*. This subset of patients might benefit from aggressive adjuvant therapy following surgical resection of the primary tumor.

15. CONSIDERATIONS FOR THE APPLICATION OF MOLECULAR TESTING IN CANCER SCREENING

Because of the public health threat cancer poses, investigators have been searching for clues to better understand carcinogenesis. The "War on Cancer" declared over 30 yr ago by the Nixon Administration is far from over and many believe it has only just begun. Although numerous genes have been identified that are associated with both sporadic and familial tumors, the diagnostic and prognostic relevance of many of these remains to be determined. The fact that cancer itself arises as a result of many different alterations to the genome presents a diagnostic dilemma for the clinical laboratory. Furthermore, the inconsistent presence of specific gene mutations in individual tumors of a single type represents a formidable problem to the application of molecular diagnostics in cancer screening.

For familial cancers, family members can be screened for predisposing risks if identification of a mutant gene or marker sequence associated with that type of tumor has been established in the germline of the affected individual. Even in these cases, which seem to be relatively straightforward, the data must be interpreted and handled very cautiously. As we have learned from the search for mutations in the *BRCA1* gene in familial cases of breast cancer, this type of laboratory testing has opened a Pandora's box with respect to ethical, legal, and technical issues never before imagined. Determining a patient's predisposition to development of a sporadic tumor and the associated diagnostic/prognostic issues will continue to challenge

the laboratory in an unprecedented fashion. As the definitive diagnosis of cancer is dependent on histological evaluation of tissues by a pathologist, the College of American Pathologists has published a summary of clinically relevant prognostic markers in solid tumors *(360–363)*. The general conclusions for solid tumors such as breast, colon, and prostate cancers supported the fact that many biological markers are currently available. However, with respect to many of them, considerable technical advances need to be made for routine analysis in the clinical laboratory and clinical data must evaluated to determine the relevance of any particular marker to patient management. As an evolving area of clinical research, future studies should identify new markers for prognostication and predisposition testing as well as confirm the efficacy of existing potential markers.

16. CONCLUSIONS

A large amount of evidence has now accumulated suggesting a genetic basis for the development of neoplastic disease in humans. However, the genetic damage that has been documented in various human cancers includes both large-scale alterations (chromosomal aberrations and ploidy changes) and DNA sequence alterations (single nucleotide changes or alterations in short segments of DNA). In addition, the patterns of genetic damage within a single tumor can vary from a few molecular alterations at specific loci to genomewide mutations involving a large number of loci. Several distinct forms of genomic instability might provide the molecular basis for neoplastic transformation in humans. Cells undergoing neoplastic transformation might accumulate genetic damage related to progressive genomic instability or the result of episodic (or transient) genomic instability. Transforming mutations could arise through either of these mechanisms, involving chromosomal alterations or sequence alterations (point mutations and/or microsatellite instability). Although the significance of mutations to the etiological mechanisms of tumor development has been debated, a causal role for genetic lesions in the genesis of cancer is commonly accepted. Thus, genetic lesions represent an integral part of the processes of neoplastic transformation, tumorigenesis, and tumor progression, and as such, they represent potentially valuable markers for cancer detection, diagnosis, staging, and prediction of clinical outcome *(6,7)*. Through the application of specific and sensitive molecular methodologies, the molecular diagnostic laboratory of the future will be able to effectively screen populations at high risk for the development of cancer, potentially impacting the early detection and diagnosis of human cancers. In addition, development of new molecular diagnostic assays will expand the ability of clinicians to accurately stage tumor development, monitor progression of metastatic disease, and evaluate therapeutic outcome, facilitating the application of effective intervention strategies in the treatment of human tumors.

REFERENCES

1. Foulds, L. The natural history of cancer. *J. Chronic Dis.* **8:**2–37, 1958.
2. Weinberg, R. A. Oncogenes, antioncogenes, and the molecular bases of multistep carcinogenesis. *Cancer Res.* **49:**3713–3721, 1989.
3. Bishop, J. M. Molecular themes in oncogenesis. *Cell* **64:**235–248, 1991.
4. Vogelstein, B., Fearon, E. R., Hamilton, S. R., et al. Genetic alterations during colorectal-tumor development. *N. Engl. J. Med.* **319:** 525–532, 1988.
5. Fearon, E. R. and Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61:**759–767, 1990.
6. Mao, L. and Sidransky, D. Cancer screening based on genetic alterations in human tumors. *Cancer Res.* **54:**1939s–1940s, 1994.
7. Sidransky, D. Molecular markers in cancer: can we make better predictions? *Int. J. Cancer* **64:**1–2, 1995.
8. Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194:**23–28, 1976.
9. Lengauer, C., Kinzler, K. W., and Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396:**643–649, 1998.
10. Renan, M. J. How many mutations are required for tumorigenesis? Implications from human cancer data. *Mol. Carcinog.* **7:**139–146, 1993.
11. Cairns, J. Mutation selection and the natural history of cancer. *Nature* **255:**197–200, 1975.
12. Loeb, K. R. and Loeb, L. A. Significance of multiple mutations in cancer. *Carcinogenesis* **21:**379–385, 2000.
13. Cohen, A. M., Minsky, B. D., and Schilsky, R. L. Cancer of the colon, in *Cancer: Principles and Practice of Oncology,* 5th ed., DeVita, V. T., Hellman, S., and Rosenberg, S. A., eds., Lippincott–Raven, Philadelphia, 1997, pp. 1144–1197.
14. Kinzler, K. W. and Vogelstein, B. Colorectal tumors, in *The Metabolic and Molecular Bases of Inherited Disease,* 8th ed., Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds., McGraw-Hill, New York, 2001, pp. 1033–1062.
15. Sidransky, D. Molecular markers in cancer diagnosis. *J. Natl. Cancer Inst. Monogr.* **17:**27–29, 1995.
16. Crow, J. F. The high spontaneous mutation rate: is it a health risk? *Proc. Natl. Acad. Sci. USA* **94:**8380–8386, 1997.
17. Cairns, J. Mutation and cancer: the antecedents to our studies of adaptive mutation. *Genetics* **148:**1433–1440, 1998.
18. Ruddon, R. W. *Cancer Biology,* 3rd ed., Oxford University Press, New York, 1995.
19. Fisher, J. C. Multiple-mutation theory of carcinogenesis. *Nature* **181:**651–652, 1958.
20. Fisher, J. C. and Hollomon, J. H. A hypothesis for the origin of cancer foci. *Cancer* **4:**916–918, 1951.
21. Strong, L. C. The induction of mutations by a carcinogen. *Br. J. Cancer* **3:**97–108, 1949.
22. Berenblum, I. and Shubik, P. An experimental study of the initiating stage of carcinogenesis, and re-examination of the somatic cell mutation theory of cancer. *Br. J. Cancer* **3:**109–118, 1949.
23. Nowell, P. C. and Hungerford, D. A. Chromosome studies on normal and leukemic human leukocytes. *J. Natl. Cancer Inst.* **25:**85–109, 1960.
24. Rowley, J. D. A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. [letter]. *Nature* **243:**290–293, 1973.
25. Albertson, D. G., Collins, C., McCormick, F., and Gray, J. W. Chromosome aberrations in solid tumors. *Nat. Genet.* **34:**369–376, 2003.
26. Hussain, S. P. and Harris, C. C. Molecular epidemiology of human cancer: contribution of mutation spectra studies of tumor suppressor genes. *Cancer Res.* **58:**4023–4037, 1998.
27. Perou, C. M., Sorlie, T., Eisen, M. B., et al. Molecular portraits of human breast tumours. *Nature* **406:**747–752, 2000.
28. Hedenfalk, I., Ringner, M., Ben-Dor, A., et al. Molecular classification of familial non-BRCA1/BRCA2 breast cancer. *Proc. Natl. Acad. Sci. USA* **100:**2532–2537, 2003.
29. Chung, C. H., Parker, J. S., Karaca, G., et al. Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. *Cancer Cell* **5:**489–500, 2004.
30. Meyerson, M., Franklin, W. A., and Kelley, M. J. Molecular classification and molecular genetics of human lung cancers. *Semin. Oncol.* **31:**4–19, 2004.

31. Hoang, C. D., D'Cunha, J., Tawfic, S. H., Gruessner, A. C., Kratzke, R. A., and Maddaus, M. A. Expression profiling of non-small cell lung carcinoma identifies metastatic genotypes based on lymph node tumor burden. *J. Thorac. Cardiovasc. Surg.* **127:**1332–1341, 2004; discussion 1342.
32. Warner, G. C., Reis, P. P., Jurisica, I., et al. Molecular classification of oral cancer by cDNA microarrays identifies overexpressed genes correlated with nodal metastasis. *Int. J. Cancer* **110:**857–868, 2004.
33. Cleator, S. and Ashworth, A. Molecular profiling of breast cancer: clinical implications. *Br. J. Cancer* **90:**1120–1124, 2004.
34. Troester, M. A., Hoadley, K. A., Sorlie, T., et al. Cell-type-specific responses to chemotherapeutics in breast cancer. *Cancer Res.* **64:**4218–4226, 2004.
35. Armitage, P. and Doll, R. A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. *Br. J. Cancer* **11:**161–169, 1957.
36. Knudson, A. G., Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. USA* **68:**820–823, 1971.
37. Baylin, S. B. Mechanisms underlying epigenetically mediated gene silencing in cancer. *Semin. Cancer Biol.* **12:**331–337, 2002.
38. Herman, J. G. and Baylin, S. B. Gene silencing in cancer in association with promoter hypermethylation. *N. Engl. J. Med.* **349:** 2042–2054, 2003.
39. Fazzari, M. J. and Greally, J. M. Epigenomics: beyond CpG islands. *Nat. Rev. Genet.* **5:**446–455, 2004.
40. Jackson, A. L. and Loeb, L. A. The mutation rate and cancer. *Genetics* **148:**1483–1493, 1998.
41. Loeb, L. A. A mutator phenotype in cancer. *Cancer Res.* **61:** 3230–3239, 2001.
42. Loeb, L. A. and Christians, F. C. Multiple mutations in human cancers. *Mutat. Res.* **350:**279–286, 1996.
43. Perucho, M. Cancer of the microsatellite mutator phenotype. *Biol. Chem.* **377:**675–684, 1996.
44. Orr-Weaver, T. L. and Weinberg, R. A. A checkpoint on the road to cancer. *Nature* **392:**223–224, 1998.
45. Christians, F. C., Newcomb, T. G., and Loeb, L. A. Potential sources of multiple mutations in human cancers. *Prev. Med.* **24:**329–332, 1995.
46. Friedberg, E. C., McDaniel, L. D., and Schultz, R. A. The role of endogenous and exogenous DNA damage and mutagenesis. *Curr. Opin. Genet. Dev.* **14:**5–10, 2004.
47. Sancar, A., Lindsey-Boltz, L. A., Unsal-Kaccmaz, K., and Linn, S. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu. Rev. Biochem.* **73:**39–85, 2004.
48. Cooper, D. N., and Youssoufian, H. The CpG dinucleotide and human genetic disease. *Hum. Genet.* **78:**151–155, 1988.
49. Goyette, M. C., Cho, K., Fasching, C. L., et al. Progression of colorectal cancer is associated with multiple tumor suppressor gene defects but inhibition of tumorigenicity is accomplished by correction of any single defect via chromosome transfer. *Mol. Cell. Biol.* **12:**1387–1395, 1992.
50. Duesberg, P., Rausch, C., Rasnick, D., and Hehlmann, R. Genetic instability of cancer cells is proportional to their degree of aneuploidy. *Proc. Natl. Acad. Sci. USA* **95:**13,692–13,697, 1998.
51. Strauss, B. S. Hypermutability in carcinogenesis. *Genetics* **148:** 1619–1626, 1998.
52. Sieber, O. M., Heinimann, K., and Tomlinson, I. P. Genomic instability—the engine of tumorigenesis? *Nat. Rev. Cancer* **3:** 701–708, 2003.
53. Hollstein, M., Shomer, B., Greenblatt, M., et al. Somatic point mutations in the p53 gene of human tumors and cell lines: updated compilation. *Nucleic Acids Res.* **24:**141–146, 1996.
54. Moolgavkar, S. H. and Luebeck, E. G. Multistage carcinogenesis and the incidence of human cancer. *Genes Chromosomes Cancer* **38:**302–306, 2003.
55. Boesen, J. J., Niericker, M. J., Dieteren, N., and Simons, J. W. How variable is a spontaneous mutation rate in cultured mammalian cells? *Mutat. Res.* **307:**121–129, 1994.
56. Eldridge, S. R. and Gould, M. N. Comparison of spontaneous mutagenesis in early-passage human mammary cells from normal and malignant tissues. *Int. J. Cancer* **50:**321–324, 1992.
57. Wittenkeller, J. L., Storer, B., Bittner, G., and Schiller, J. H. Comparison of spontaneous and induced mutation rates in an immortalized human bronchial epithelial cell line and its tumorigenic derivative. *Oncology* **54:**335–341, 1997.
58. Loeb, L. A. Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res.* **51:**3075–3079, 1991.
59. Loeb, L. A. Transient expression of a mutator phenotype in cancer cells. *Science* **277:**1449–1450, 1997.
60. Richards, B., Zhang, H., Phear, G., and Meuth, M. Conditional mutator phenotypes in hMSH2-deficient tumor cell lines. *Science* **277:**1523–1526, 1997.
61. Bhattacharyya, N. P., Skandalis, A., Ganesh, A., Groden, J., and Meuth, M. Mutator phenotypes in human colorectal carcinoma cell lines. *Proc. Natl. Acad. Sci. USA* **91:**6319–6323, 1994.
62. Eshleman, J. R., Lang, E. Z., Bowerfind, G. K., et al. Increased mutation rate at the hprt locus accompanies microsatellite instability in colon cancer. *Oncogene* **10:**33–37, 1995.
63. Glaab, W. E. and Tindall, K. R. Mutation rate at the hprt locus in human cancer cell lines with specific mismatch repair-gene defects. *Carcinogenesis* **18:**1–8, 1997.
64. Tlsty, T. D., Margolin, B. H., and Lum, K. Differences in the rates of gene amplification in nontumorigenic and tumorigenic cell lines as measured by Luria–Delbruck fluctuation analysis. *Proc. Natl. Acad. Sci. USA* **86:**9441–9445, 1989.
65. Eshleman, J. R. and Markowitz, S. D. Mismatch repair defects in human carcinogenesis. *Hum. Mol. Genet.* **5:**1489–1494, 1996.
66. Lengauer, C., Kinzler, K. W., and Vogelstein, B. Genetic instability in colorectal cancers. *Nature* **386:**623–627, 1997.
67. Cahill, D. P., Lengauer, C., Yu, J., et al. Mutations of mitotic checkpoint genes in human cancers. *Nature* **392:**300–303, 1998.
68. Greenblatt, M. S., Bennett, W. P., Hollstein, M., and Harris, C. C. Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. *Cancer Res.* **54:**4855–4878, 1994.
69. Baker, S. J., Fearon, E. R., Nigro, J. M., et al. Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science* **244:**217–221, 1989.
70. Bos, J. L. ras oncogenes in human cancer: a review. *Cancer Res.* **49:**4682–4689, 1989.
71. Dalla-Favera, R., Martinotti, S., Gallo, R. C., Erikson, J., and Croce, C. M. Translocation and rearrangements of the c-myc oncogene locus in human undifferentiated B-cell lymphomas. *Science* **219:** 963–967, 1983.
72. Coleman, W. B. and Tsongalis, G. J. The role of genomic instability in the development of human cancer, in *The Molecular Basis of Human Cancer,* Coleman, W. B. and Tsongalis, G. J., eds., Humana Pr, Totowa, NJ, 2002, pp. 115–142.
73. Rosenberg, S. M., Thulin, C., and Harris, R. S. Transient and heritable mutators in adaptive evolution in the lab and in nature. *Genetics* **148:**1559–1566, 1998.
74. Aaltonen, L. A., Peltomaki, P., Leach, F. S., et al. Clues to the pathogenesis of familial colorectal cancer. *Science* **260:**812–816, 1993.
75. Aaltonen, L. A., Peltomaki, P., Mecklin, J. P., et al. Replication errors in benign and malignant tumors from hereditary nonpolyposis colorectal cancer patients. *Cancer Res.* **54:**1645–1648, 1994.
76. Akiyama, Y., Iwanaga, R., Saitoh, K., et al. Transforming growth factor beta type II receptor gene mutations in adenomas from hereditary nonpolyposis colorectal cancer. *Gastroenterology* **112:**33–39, 1997.
77. Feig, D. I., Reid, T. M., and Loeb, L. A. Reactive oxygen species in tumorigenesis. *Cancer Res.* **54:**1890s–1894s, 1994.
78. Jackson, A. L., Chen, R., and Loeb, L. A. Induction of microsatellite instability by oxidative DNA damage. *Proc. Natl. Acad. Sci. USA* **95:**12,468–12,473, 1998.
79. Baba, S. Recent advances in molecular genetics of colorectal cancer. *World J. Surg.* **21:**678–687, 1997.
80. Yamada, N. A., Parker, J. M., and Farber, R. A. Mutation frequency analysis of mononucleotide and dinucleotide repeats after oxidative stress. *Environ. Mol. Mutagen.* **42:**75–84, 2003.
81. Sarasin, A. An overview of the mechanisms of mutagenesis and carcinogenesis. *Mutat. Res.* **544:**99–106, 2003.

82. Kastan, M. B. and Kuerbitz, S. J. Control of G1 arrest after DNA damage. *Environ. Health Perspect.* **101(Suppl. 5):**55–58, 1993.
83. Coleman, W. B. and Tsongalis, G. J. Multiple mechanisms account for genomic instability and molecular mutation in neoplastic transformation. *Clin. Chem.* **41:**644–657, 1995.
84. Marti, T. M. and Fleck, O. DNA repair nucleases. *Cell. Mol. Life Sci* **61:**336–354, 2004.
85. Bootsma, D., Kraemer, K. H., Cleaver, J. E., and Hoeijmakers, J. H. J. Nucleotide excision repair syndromes: Xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy, in *The Metabolic and Molecular Basis of Inherited Disease,* 8th ed., Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds., McGraw-Hill, New York, 2001, pp. 677–704.
86. Gatti, R. A. Ataxia–telangiectasia, in *The Metabolic and Molecular Bases of Inherited Disease,* 8th ed., Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds., McGraw-Hill, New York, 2001, pp. 705–732.
87. Swift, M., Morrell, D., Massey, R. B., and Chase, C. L. Incidence of cancer in 161 families affected by ataxia–telangiectasia. *N. Engl. J. Med.* **325:**1831–1836, 1991.
88. Thorstenson, Y. R., Roxas, A., Kroiss, R., et al. Contributions of ATM mutations to familial breast and ovarian cancer. *Cancer Res.* **63:**3325–3333, 2003.
89. Auerbach, A. D., Buchwald, M., and Joenje, H. Fanconi anemia, in *The Metabolic and Molecular Bases of Inherited Disease,* 8th ed., Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds., McGraw-Hill, New York, 2001, pp. 753–768.
90. Bagby, G. C., Jr. Genetic basis of Fanconi anemia. *Curr. Opin. Hematol.* **10:**68–76, 2003.
91. German, J. Bloom's syndrome. XX. The first 100 cancers. *Cancer Genet. Cytogenet.* **93:**100–106, 1997.
92. German, J. and Ellis, N. A. Bloom syndrome, in *The Metabolic and Molecular Basis of Inherited Disease,* 8th ed., Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds., McGraw-Hill, New York, 2001, pp. 733–752.
93. Langlois, R. G., Bigbee, W. L., Jensen, R. H., and German, J. Evidence for increased in vivo mutation and somatic recombination in Bloom's syndrome. *Proc. Natl. Acad. Sci. USA* **86:**670–674, 1989.
94. Norgauer, J., Idzko, M., Panther, E., Hellstern, O., and Herouy, Y. Xeroderma pigmentosum. *Eur. J. Dermatol.* **13:**4–9, 2003.
95. Gatti, R. A. Localizing the genes for ataxia-telangiectasia: a human model for inherited cancer susceptibility. *Adv. Cancer Res.* **56:** 77–104, 1991.
96. Digweed, M. Human genetic instability syndromes: single gene defects with increased risk of cancer. *Toxicol. Lett.* **67:**259–281, 1993.
97. Groden, J., and German, J. Bloom's syndrome. XVIII. Hypermutability at a tandem-repeat locus. *Hum. Genet.* **90:**360–367, 1992.
98. Ellis, N. A., Groden, J., Ye, T. Z., et al. The Bloom's syndrome gene product is homologous to RecQ helicases. *Cell* **83:**655–666, 1995.
99. Cleaver, J. E., and Kraemer, K. H. Xeroderma pigmentosum and Cockayne syndrome, in *The Metabolic and Molecualr Bases of Inherited Disease,* 7th ed., Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds., McGraw-Hill, New York, 1995, pp. 4393–4419.
100. Savitsky, K., Bar-Shira, A., Gilad, S., et al. A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* **268:**1749–1753, 1995.
101. Shiloh, Y. ATM and related protein kinases: safeguarding genome integrity. *Nat. Rev. Cancer* **3:**155–168, 2003.
102. Mitelman, F. *Catalog of Chromosome Aberrations in Cancer,* 5th ed, Wiley–Liss, New York, 1994.
103. Meltzer, P. S., Kallioniemi, A., and Trent, J. M. Chromosome alterations in human solid tumors, in *The Metabolic and Molecular Bases of Inhierited Disease,* 8th ed., Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds., McGraw-Hill, New York, 2001, pp. 575–596.
104. Cahill, D. P. and Lengauer, C. Tumor genome instabilities, in *The Metabolic and Molecular Bases of Inherited Disease,* 8th ed., Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds., McGraw-Hill, New York, 2001, pp. 611–612.
105. Wang, Z., Cummins, J. M., Shen, D., et al. Three classes of genes mutated in colorectal cancers with chromosomal instability. *Cancer Res.* **64:**2998–3001, 2004.
106. Vogelstein, B., Fearon, E. R., Kern, S. E., et al. Allelotype of colorectal carcinomas. *Science* **244:**207–211,1989.
107. Tsuchiya, E., Nakamura, Y., Weng, S. Y., et al. Allelotype of non-small cell lung carcinoma—comparison between loss of heterozygosity in squamous cell carcinoma and adenocarcinoma. *Cancer Res.* **52:**2478–2481, 1992.
108. Seymour, A. B., Hruban, R. H., Redston, M., et al. Allelotype of pancreatic adenocarcinoma. *Cancer Res.* **54:**2761–2764, 1994.
109. Roncalli, M., Borzio, M., Bianchi, P., and Laghi, L. Comprehensive allelotype study of hepatocellular carcinoma. *Hepatology* **32:**876, 2000.
110. Miller, B. J., Wang, D., Krahe, R., and Wright, F. A. Pooled analysis of loss of heterozygosity in breast cancer: a genome scan provides comparative evidence for multiple tumor suppressors and identifies novel candidate regions. *Am. J. Hum. Genet.* **73:**748–767, 2003.
111. Thiagalingam, S., Laken, S., Willson, J. K., et al. Mechanisms underlying losses of heterozygosity in human colorectal cancers. *Proc. Natl. Acad. Sci. USA* **98:**2698–2702, 2001.
112. El-Naggar, A. K. and Vielh, P. Solid tumor DNA content analysis. *Methods Mol. Biol.* **263:**355–370, 2004.
113. Kallioniemi, O. P., Kallioniemi, A., Piper, J., et al. Optimizing comparative genomic hybridization for analysis of DNA sequence copy number changes in solid tumors. *Genes Chromosomes Cancer* **10:** 231–243, 1994.
114. Kallioniemi, O. P., Kallioniemi, A., Sudar, D., et al. Comparative genomic hybridization: a rapid new method for detecting and mapping DNA amplification in tumors. *Semin. Cancer Biol.* **4:**41–46, 1993.
115. Bayani, J. M. and Squire, J. A. Applications of SKY in cancer cytogenetics. *Cancer Invest.* **20:**373–386, 2002.
116. Kakazu, N., Bar-Am, I., Hada, S., Ago, H., and Abe, T. A new chromosome banding technique, spectral color banding (SCAN), for full characterization of chromosomal abnormalities. *Genes Chromosomes Cancer* **37:**412–416, 2003.
117. Solomon, E., Borrow, J., and Goddard, A. D. Chromosome aberrations and cancer. *Science* **254:**1153–1160, 1991.
118. Patel, A. S., Hawkins, A. L., and Griffin, C. A. Cytogenetics and cancer. *Curr. Opin. Oncol.* **12:**62–67, 2000.
119. Radford, D. M., Fair, K. L., Phillips, N. J., et al. Allelotyping of ductal carcinoma in situ of the breast: deletion of loci on 8p, 13q, 16q, 17p and 17q. *Cancer Res.* **55:**3399–3405, 1995.
120. Boige, V., Laurent-Puig, P., Fouchet, P., et al. Concerted nonsyntenic allelic losses in hyperploid hepatocellular carcinoma as determined by a high-resolution allelotype. *Cancer Res.* **57:**1986–1990, 1997.
121. Nowak, M. A., Komarova, N. L., Sengupta, A., et al. The role of chromosomal instability in tumor initiation. *Proc. Natl. Acad. Sci. USA* **99:**16,226–16,231, 2002.
122. Rajagopalan, H., Nowak, M. A., Vogelstein, B., and Lengauer, C. The significance of unstable chromosomes in colorectal cancer. *Nat. Rev. Cancer* **3:**695–701, 2003.
123. Ghadimi, B. M., Sackett, D. L., Difilippantonio, M. J., et al. Centrosome amplification and instability occurs exclusively in aneuploid, but not in diploid colorectal cancer cell lines, and correlates with numerical chromosomal aberrations. *Genes Chromosomes Cancer* **27:**183–190, 2000.
124. Harwood, J., Tachibana, A., Davis, R., Bhattacharyya, N. P., and Meuth, M. High rate of multilocus deletion in a human tumor cell line. *Hum. Mol. Genet.* **2:**165–171, 1993.
125. Phear, G., Bhattacharyya, N. P., and Meuth, M. Loss of heterozygosity and base substitution at the APRT locus in mismatch-repair-proficient and -deficient colorectal carcinoma cell lines. *Mol. Cell. Biol.* **16:**6516–6523, 1996.
126. Le Beau, M. M. Molecular biology of cancer: cytogenetics, in *Cancer: Principles and Practice of Oncology,* 5th ed., DeVita, V. T., Hellman, S., and Rosenberg, S. A., eds., Lippincott–Raven, Philadelphia, PA, pp. 103–119.

127. Gollin, S. M. Chromosomal instability. *Curr. Opin. Oncol.* **16:** 25–31, 2004.
128. Kastan, M. B., Onyekwere, O., Sidransky, D., Vogelstein, B., and Craig, R. W. Participation of p53 protein in the cellular response to DNA damage. *Cancer Res.* **51:**6304-6311, 1991.
129. Jong, Y. J., Li, L. H., Tsou, M. H., et al. Chromosomal comparative genomic hybridization abnormalities in early- and late-onset human breast cancers: correlation with disease progression and TP53 mutations. *Cancer Genet. Cytogenet.* **148:**55–65, 2004.
130. Sugai, T., Takahashi, H., Habano, W., et al. Analysis of genetic alterations, classified according to their DNA ploidy pattern, in the progression of colorectal adenomas and early colorectal carcinomas. *J. Pathol.* **200:**168–176, 2003.
131. Filatov, L., Golubovskaya, V., Hurt, J. C., Byrd, L. L., Phillips, J. M., and Kaufmann, W. K. Chromosomal instability is correlated with telomere erosion and inactivation of G2 checkpoint function in human fibroblasts expressing human papillomavirus type 16 E6 oncoprotein. *Oncogene* **16:**1825–1838, 1998.
132. Honma, M., Momose, M., Tanabe, H., et al. Requirement of wild-type p53 protein for maintenance of chromosomal integrity. *Mol. Carcinog.* **28:**203–214, 2000.
133. Tarapore, P. and Fukasawa, K. p53 mutation and mitotic infidelity. *Cancer Invest.* **18:**148–155, 2000.
134. Shih, I. M., Zhou, W., Goodman, S. N., Lengauer, C., Kinzler, K. W., and Vogelstein, B. Evidence that genetic instability occurs at an early stage of colorectal tumorigenesis. *Cancer Res.* **61:**818–822, 2001.
135. Baker, S. J., Preisinger, A. C., Jessup, J. M., et al. p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis. *Cancer Res.* **50:**7717–7722, 1990.
136. Eshleman, J. R., Casey, G., Kochera, M. E., et al. Chromosome number and structure both are markedly stable in RER colorectal cancers and are not destabilized by mutation of p53. *Oncogene* **17:**719–725, 1998.
137. Kramer, A., Neben, K., and Ho, A. D. Centrosome replication, genomic instability and cancer. *Leukemia* **16:**767–775, 2002.
138. Lingle, W. L., Barrett, S. L., Negron, V. C., et al. Centrosome amplification drives chromosomal instability in breast tumor development. *Proc. Natl. Acad. Sci. USA* **99:**1978–1983, 2002.
139. Bischoff, J. R., Anderson, L., Zhu, Y., et al. A homologue of *Drosophila aurora* kinase is oncogenic and amplified in human colorectal cancers. EMBO *J.* **17:**3052–3065.
140. Zhou, H., Kuang, J., Zhong, L., et al. Tumour amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. *Nat. Genet.* **20:**189–193, 1998.
141. Wolf, G., Elez, R., Doermer, A., et al. Prognostic significance of polo-like kinase (PLK) expression in non-small cell lung cancer. *Oncogene* **14:**543–549, 1997.
142. Tanaka, T., Kimura, M., Matsunaga, K., Fukada, D., Mori, H., and Okano, Y. Centrosomal kinase AIK1 is overexpressed in invasive ductal carcinoma of the breast. *Cancer Res.* **59:**2041–2044, 1999.
143. Carroll, P. E., Okuda, M., Horn, H. F., et al. Centrosome hyperamplification in human cancer: chromosome instability induced by p53 mutation and/or Mdm2 overexpression. *Oncogene* **18:**1935–1944, 1999.
144. Tarapore, P. and Fukasawa, K. Loss of p53 and centrosome hyperamplification. *Oncogene* **21:**6234–6240, 2002.
145. Tutt, A., Gabriel, A., Bertwistle, D., et al. Absence of Brca2 causes genome instability by chromosome breakage and loss associated with centrosome amplification. *Curr. Biol.* **9:**1107–1110, 1999.
146. Deng, C. X. Roles of BRCA1 in centrosome duplication. *Oncogene* **21:**6222–6227, 2002.
147. Bharadwaj, R. and Yu, H. The spindle checkpoint, aneuploidy, and cancer. *Oncogene* **23:**2016–2027, 2004.
148. Hartwell, L. H. and Kastan, M. B. Cell cycle control and cancer. *Science* **266:**1821–1828, 1994.
149. Cahill, D. P., da Costa, L. T., Carson-Walter, E. B., Kinzler, K. W., Vogelstein, B., and Lengauer, C. Characterization of MAD2B and other mitotic spindle checkpoint genes. *Genomics* **58:**181–187, 1999.
150. Li, Y. and Benezra, R. Identification of a human mitotic checkpoint gene: hsMAD2. *Science* **274:**246–248, 1996.
151. Jaffrey, R. G., Pritchard, S. C., Clark, C., et al. Genomic instability at the BUB1 locus in colorectal cancer, but not in non-small cell lung cancer. *Cancer Res.* **60:**4349–4352, 2000.
152. Myrie, K. A., Percy, M. J., Azim, J. N., Neeley, C. K., and Petty, E. M. Mutation and expression analysis of human BUB1 and BUB1B in aneuploid breast cancer cell lines. *Cancer Lett.* **152:**193–199, 2000.
153. Sato, M., Sekido, Y., Horio, Y., et al. Infrequent mutation of the hBUB1 and hBUBR1 genes in human lung cancer. *Jpn. J. Cancer Res.* **91:**504–509, 2000.
154. Yamaguchi, K., Okami, K., Hibi, K., Wehage, S. L., Jen, J., and Sidransky, D. Mutation analysis of hBUB1 in aneuploid HNSCC and lung cancer cell lines. *Cancer Lett.* **139:**183–187, 1999.
155. Imai, Y., Shiratori, Y., Kato, N., Inoue, T., and Omata, M. Mutational inactivation of mitotic checkpoint genes, hsMAD2 and hBUB1, is rare in sporadic digestive tract cancers. *Jpn. J. Cancer Res.* **90:** 837–840, 1999.
156. Gualberto, A., Aldape, K., Kozakiewicz, K., and Tlsty, T. D. An oncogenic form of p53 confers a dominant, gain-of-function phenotype that disrupts spindle checkpoint control. *Proc. Natl. Acad. Sci. USA* **95:**5166–5171, 1998.
157. Shigeta, T., Takagi, M., Delia, D., et al. Defective control of apoptosis and mitotic spindle checkpoint in heterozygous carriers of ATM mutations. *Cancer Res.* **59:**2602–2607, 1999.
158. Canman, C. E. and Lim, D. S. The role of ATM in DNA damage responses and cancer. *Oncogene* **17:**3301–3308, 1998.
159. Zhang, H., Tombline, G., and Weber, B. L. BRCA1, BRCA2, and DNA damage response: collision or collusion? *Cell* **92:**433–436, 1998.
160. Tlsty, T. D., Briot, A., Gualberto, A., et al. Genomic instability and cancer. *Mutat. Res.* **337:**1–7, 1995.
161. Hogarty, M. D. and Brodeur, G. M. Gene amplification in human cancers: biological and clinical significance, in *The Metabolic and Molecular Bases of Inherited Disease,* 8th ed., Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds., McGraw-Hill, New York, 2001, pp. 597–610.
162. Livingstone, L. R., White, A., Sprouse, J., Livanos, E., Jacks, T., and Tlsty, T. D. Altered cell cycle arrest and gene amplification potential accompany loss of wild-type p53. *Cell* **70:**923–935, 1992.
163. Yin, Y., Tainsky, M. A., Bischoff, F. Z., Strong, L. C., and Wahl, G. M. Wild-type p53 restores cell cycle control and inhibits gene amplification in cells with mutant p53 alleles. *Cell* **70:**937–948, 1992.
164. Oren, M. Relationship of p53 to the control of apoptotic cell death. *Semin. Cancer Biol.* **5:**221–227, 1994.
165. Dubeau, L. Ovarian cancer, in *The Metabolic and Molecular Bases of Inherited Disease,* 8th ed., Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds., McGraw-Hill, New York, 2001, pp. 1091–1096.
166. Hruban, R. H., Wilentz, R. E., and Kern, S. E. Genetic progression in the pancreatic ducts. *Am. J. Pathol.* **156:**1821–1825, 2000.
167. Croce, C. M. Chromosome translocations and human cancer. *Cancer Res.* **46:**6019–6023, 1986.
168. Groden, J., Thliveris, A., Samowitz, W., et al. Identification and characterization of the familial adenomatous polyposis coli gene. *Cell* **66:**589–600, 1991.
169. Kinzler, K. W., Nilbert, M. C., Su, L. K., et al. Identification of FAP locus genes from chromosome 5q21. *Science* **253:**661–665, 1991.
170. Kinzler, K. W., Nilbert, M. C., Vogelstein, B., et al. Identification of a gene located at chromosome 5q21 that is mutated in colorectal cancers. *Science* **251:**1366–1370, 1991.
171. Nishisho, I., Nakamura, Y., Miyoshi, Y., et al. Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. *Science* **253:**665–669, 1991.
172. Fearon, E. R., Cho, K. R., Nigro, J. M., et al. Identification of a chromosome 18q gene that is altered in colorectal cancers. *Science* **247:**49–56, 1990.
173. Sutherland, G. R. and Richards, R. I. Simple tandem DNA repeats and human genetic disease. *Proc. Natl. Acad. Sci. USA* **92:**3636–3641, 1995.
174. Peinado, M. A., Malkhosyan, S., Velazquez, A., and Perucho, M. Isolation and characterization of allelic losses and gains in colorectal tumors by arbitrarily primed polymerase chain reaction. *Proc. Natl. Acad. Sci. USA* **89:**10,065–10,069, 1992.

175. Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D., and Perucho, M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* **363:** 558–561, 1993.
176. Thibodeau, S. N., Bren, G., and Schaid, D. Microsatellite instability in cancer of the proximal colon. *Science* **260:**816–819, 1993.
177. Zhang, L., Yu, J., Willson, J. K., Markowitz, S. D., Kinzler, K. W., and Vogelstein, B. Short mononucleotide repeat sequence variability in mismatch repair-deficient cancers. *Cancer Res.* **61:** 3801–3805, 2001.
178. Dietmaier, W., Wallinger, S., Bocker, T., Kullmann, F., Fishel, R., and Ruschoff, J. Diagnostic microsatellite instability: definition and correlation with mismatch repair protein expression. *Cancer Res.* **57:**4749–4756, 1997.
179. Thibodeau, S. N., French, A. J., Cunningham, J. M., et al. Microsatellite instability in colorectal cancer: different mutator phenotypes and the principal involvement of hMLH1. *Cancer Res.* **58:** 1713–1718, 1998.
180. Frazier, M. L., Sinicrope, F. A., Amos, C. I., et al. Loci for efficient detection of microsatellite instability in hereditary non-polyposis colorectal cancer. *Oncol. Rep.* **6:**497–505, 1999.
181. Yamada, N. A., Castro, A., and Farber, R. A. Variation in the extent of microsatellite instability in human cell lines with defects in different mismatch repair genes. *Mutagenesis* **18:**277–282, 2003.
182. Honchel, R., Halling, K. C., and Thibodeau, S. N. Genomic instability in neoplasia. *Semin. Cell. Biol.* **6:**45–52, 1995.
183. Peltomaki, P. DNA mismatch repair gene mutations in human cancer. *Environ. Health Perspect.* **105 (Suppl. 4):**775–780, 1997.
184. Hoang, J. M., Cottu, P. H., Thuille, B., Salmon, R. J., Thomas, G., and Hamelin, R. BAT-26, an indicator of the replication error phenotype in colorectal cancers and cell lines. *Cancer Res.* **57:**300–303, 1997.
185. Zhou, X. P., Hoang, J. M., Li, Y. J., et al. Determination of the replication error phenotype in human tumors without the requirement for matching normal DNA by analysis of mononucleotide repeat microsatellites. *Genes Chromosomes Cancer* **21:**101–107, 1998.
186. Parsons, R., Myeroff, L. L., Liu, B., et al. Microsatellite instability and mutations of the transforming growth factor beta type II receptor gene in colorectal cancer. *Cancer Res.* **55:**5548–5550, 1995.
187. Parsons, R., Li, G. M., Longley, M. J., et al. Hypermutability and mismatch repair deficiency in RER+ tumor cells. *Cell* **75:**1227–1236, 1993.
188. Chen, J., Heerdt, B. G., and Augenlicht, L. H. Presence and instability of repetitive elements in sequences the altered expression of which characterizes risk for colonic cancer. *Cancer Res.* **55:**174–180, 1995.
189. Eichler, E. E., Holden, J. J., Popovich, B. W., et al. Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nat. Genet.* **8:**88–94, 1994.
190. Mao, L., Lee, D. J., Tockman, M. S., Erozan, Y. S., Askin, F., and Sidransky, D. Microsatellite alterations as clonal markers for the detection of human cancer. *Proc. Natl. Acad. Sci. USA* **91:**9871–9875, 1994.
191. Risinger, J. I., Umar, A., Barrett, J. C., and Kunkel, T. A. A hPMS2 mutant cell line is defective in strand-specific mismatch repair. *J. Biol. Chem.* **270:**18, 183–18,186, 1995.
192. Risinger, J. I., Umar, A., Boyd, J., Berchuck, A., Kunkel, T. A., and Barrett, J. C. Mutation of MSH3 in endometrial cancer and evidence for its functional role in heteroduplex repair. *Nat. Genet.* **14:**102–105, 1996.
193. Oki, E., Oda, S., Maehara, Y., and Sugimachi, K. Mutated gene-specific phenotypes of dinucleotide repeat instability in human colorectal carcinoma cell lines deficient in DNA mismatch repair. *Oncogene* **18:**2143–2147, 1999.
194. Bennett, S. E., Umar, A., Oshima, J., Monnat, R. J., Jr., and Kunkel, T. A. Mismatch repair in extracts of Werner syndrome cell lines. *Cancer Res.* **57:**2956–2960, 1997.
195. Li, G. M. and Modrich, P. Restoration of mismatch repair to nuclear extracts of H6 colorectal tumor cells by a heterodimer of human MutL homologs. *Proc. Natl. Acad. Sci. USA* **92:**1950–1954, 1995.
196. Papadopoulos, N., Nicolaides, N. C., Wei, Y. F., et al. Mutation of a mutL homolog in hereditary colon cancer. *Science* **263:**1625–1629, 1994.
197. Glaab, W. E., Risinger, J. I., Umar, A., Kunkel, T. A., Barrett, J. C., and Tindall, K. R. Characterization of distinct human endometrial carcinoma cell lines deficient in mismatch repair that originated from a single tumor. *J. Biol. Chem.* **273:**26,662–26,669, 1998.
198. Baranovskaya, S., Soto, J. L., Perucho, M., and Malkhosyan, S. R. Functional significance of concomitant inactivation of hMLH1 and hMSH6 in tumor cells of the microsatellite mutator phenotype. *Proc. Natl. Acad. Sci. USA* **98:**15107–15,112, 2001.
199. Bocker, T., Diermann, J., Friedl, W., et al. Microsatellite instability analysis: a multicenter study for reliability and quality control. *Cancer Res.* **57:**4739–4743, 1997.
200. Boland, C. R., Thibodeau, S. N., Hamilton, S. R., et al. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.* **58:**5248–5257, 1998.
201. Siah, S. P., Quinn, D. M., Bennett, G. D., et al. Microsatellite instability markers in breast cancer: a review and study showing MSI was not detected at "BAT 25" and "BAT 26" microsatellite markers in early-onset breast cancer. *Breast Cancer Res. Treat.* **60:**135–142, 2000.
202. Luqmani, Y. A. and Mathew, M., Allelic variation of BAT-25 and BAT-26 mononucleotide repeat loci in tumours from a group of young women with breast cancer. *Int. J. Oncol.* **25:**771–775, 2004.
203. Fujiwara, T., Stolker, J. M., Watanabe, T., et al. Accumulated clonal genetic alterations in familial and sporadic colorectal carcinomas with widespread instability in microsatellite sequences. *Am. J. Pathol.* **153:**1063–1078, 1998.
204. Bocker, T., Schlegel, J., Kullmann, F., et al. Genomic instability in colorectal carcinomas: comparison of different evaluation methods and their biological significance. *J. Pathol.* **179:**15–19, 1996.
205. Sood, A. K., Holmes, R., Hendrix, M. J., and Buller, R. E. Application of the National Cancer Institute international criteria for determination of microsatellite instability in ovarian cancer. *Cancer Res.* **61:**4371–4371, 2001.
206. Horii, A., Han, H. J., Shimada, M., et al. Frequent replication errors at microsatellite loci in tumors of patients with multiple primary cancers. *Cancer Res.* **54:**3373–3375, 1994.
207. Akiyama, Y., Nakasaki, H., Nihei, Z., et al. Frequent microsatellite instabilities and analyses of the related genes in familial gastric cancers. *Jpn. J. Cancer Res.* **87:**595–601, 1996.
208. Nakashima, H., Honda, M., Inoue, H., et al. Microsatellite instability in multiple gastric cancers. *Int. J. Cancer* **64:**239–242, 1995.
209. Contegiacomo, A., Palmirotta, R., De Marchis, L., et al. Microsatellite instability and pathological aspects of breast cancer. *Int. J. Cancer* **64:**264–268, 1995.
210. Paulson, T. G., Wright, F. A., Parker, B. A., Russack, V., and Wahl, G. M. Microsatellite instability correlates with reduced survival and poor disease prognosis in breast cancer. *Cancer Res.* **56:**4021–4026, 1996.
211. De Marchis, L., Contegiacomo, A., D'Amico, C., et al. Microsatellite instability is correlated with lymph node-positive breast cancer. *Clin. Cancer Res.* **3:**241–248, 1997.
212. Caldes, T., Perez-Segura, P., Tosar, A., de La Hoya, M., and Diaz-Rubio, E. Microsatellite instability correlates with negative expression of estrogen and progesterone receptors in sporadic breast cancer. *Teratog. Carcinog. Mutagen.* **20:**283–291, 2000.
213. Imyanitov, E. N., Togo, A. V., Suspitsin, E. N., et al. Evidence for microsatellite instability in bilateral breast carcinomas. *Cancer Lett.* **154:**9–17, 2000.
214. Richard, S. M., Bailliet, G., Paez, G. L., Bianchi, M. S., Peltomaki, P., and Bianchi, N. O. Nuclear and mitochondrial genome instability in human breast cancer. *Cancer Res.* **60:**4231–4237, 2000.
215. Jiricny, J., and Nystrom-Lahti, M. Mismatch repair defects in cancer. *Curr. Opin. Genet. Dev.* **10:**157–161, 2000.
216. Fishel, R., Lescoe, M. K., Rao, M. R., et al. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell.* **75:**1027–1038, 1993.

217. Leach, F. S., Nicolaides, N. C., Papadopoulos, N., et al. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* **75:**1215–1225, 1993.
218. Umar, A., Risinger, J. I., Glaab, W. E., Tindall, K. R., Barrett, J. C., and Kunkel, T. A. Functional overlap in mismatch repair by human MSH3 and MSH6. *Genetics* **148:**1637–1646, 1998.
`219. Palombo, F., Gallinari, P., Iaccarino, I., et al GTBP, a 160-kilodalton protein essential for mismatch-binding activity in human cells. *Science* **268:**1912–1914, 1995.
220. Papadopoulos, N., Nicolaides, N. C., Liu, B., et al. Mutations of GTBP in genetically unstable cells. *Science* **268:**1915–1917, 1995.
221. Bronner, C. E., Baker, S. M., Morrison, P. T., et al. Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature* **368:**258–261, 1994.
222. Nicolaides, N. C., Papadopoulos, N., Liu, B., et al. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. *Nature* **371:**75–80, 1994.
223. Lipkin, S. M., Wang, V., Jacoby, R., et al. MLH3: a DNA mismatch repair gene associated with mammalian microsatellite instability. *Nat. Genet.* **24:**27–35, 2000.
224. Bapat, B. V., Madlensky, L., Temple, L. K., et al. Family history characteristics, tumor microsatellite instability and germline MSH2 and MLH1 mutations in hereditary colorectal cancer. *Hum. Genet.* **104:**167–176, 1999.
225. Loukola, A., Vilkki, S., Singh, J., Launonen, V., and Aaltonen, L. A. Germline and somatic mutation analysis of MLH3 in MSI-positive colorectal cancer. *Am. J. Pathol.* **157:**347–152, 2000.
226. Planck, M., Wenngren, E., Borg, A., Olsson, H., and Nilbert, M. Somatic frameshift alterations in mononucleotide repeat-containing genes in different tumor types from an HNPCC family with germline MSH2 mutation. *Genes Chromosomes Cancer* **29:**33–39, 2000.
227. Muller, A. and Fishel, R. Mismatch repair and the hereditary non-polyposis colorectal cancer syndrome (HNPCC). *Cancer Invest.* **20:**102–109, 2002.
228. Liu, B., Nicolaides, N. C., Markowitz, S., et al. Mismatch repair gene defects in sporadic colorectal cancers with microsatellite instability. *Nat. Genet* **9:**48–55, 1995.
229. Moslein, G., Tester, D. J., Lindor, N. M., et al. Microsatellite instability and mutation analysis of hMSH2 and hMLH1 in patients with sporadic, familial and hereditary colorectal cancer. *Hum. Mol. Genet.* **5:**1245–1252, 1996.
230. Schmutte, C. and Fishel, R. Genomic instability: first step to carcinogenesis. *Anticancer Res.* **19:**4665–4696, 1999.
231. Arzimanoglou, II, Gilbert, F., and Barber, H. R. Microsatellite instability in human solid tumors. *Cancer* **82:**1808–1820, 1998.
232. Lothe, R. A. Microsatellite instability in human solid tumors. *Mol. Med. Today* **3:**61–68, 1997.
233. Schlotterer, C. and Tautz, D. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* **20:**211–215, 1992.
234. Coleman, W. B. and Tsongalis, G. J. The role of genomic instability in human carcinogenesis. *Anticancer Res.* **19:**4645–4664, 1999.
235. Thomas, D. C., Umar, A., and Kunkel, T. A. Microsatellite instability and mismatch repair defects in cancer. *Mutat. Res.* **350:**201–205, 1996.
236. Lovett, S. T. and Feschenko, V. V. Stabilization of diverged tandem repeats by mismatch repair: evidence for deletion formation via a misaligned replication intermediate. *Proc. Natl. Acad. Sci. USA* **93:**7120–7124, 1996.
237. Boyer, J. C. and Farber, R. A. Mutation rate of a microsatellite sequence in normal human fibroblasts. *Cancer Res.* **58:**3946–3949, 1998.
238. Hanford, M. G., Rushton, B. C., Gowen, L. C., and Farber, R. A. Microsatellite mutation rates in cancer cell lines deficient or proficient in mismatch repair. *Oncogene* **16:**2389–2393, 1998.
239. Umar, A., Koi, M., Risinger, J. I., et al. Correction of hypermutability, *N*-methyl-*N*'-nitro-*N*-nitrosoguanidine resistance, and defective DNA mismatch repair by introducing chromosome 2 into human tumor cells with mutations in MSH2 and MSH6. *Cancer Res.* **57:**3949–3955, 1997.
240. da Costa, L. T., Liu, B., el-Deiry, W., et al. Polymerase delta variants in RER colorectal tumours. *Nat. Genet.* **9:**10–11, 1995.
241. Parc, Y. R., Halling, K. C., Wang, L., et al. HMSH6 alterations in patients with microsatellite instability-low colorectal cancer. *Cancer Res.* **60:**2225–2231, 2000.
242. Verma, L., Kane, M. F., Brassett, C., et al. Mononucleotide microsatellite instability and germline MSH6 mutation analysis in early onset colorectal cancer. *J. Med. Genet.* **36:**678–682, 1999.
243. Koi, M., Umar, A., Chauhan, D. P., et al. Human chromosome 3 corrects mismatch repair deficiency and microsatellite instability and reduces *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine tolerance in colon tumor cells with homozygous hMLH1 mutation. *Cancer Res.* **54:** 4308–4312, 1994.
244. Risinger, J. I., Umar, A., Glaab, W. E., Tindall, K. R., Kunkel, T. A., and Barrett, J. C. Single gene complementation of the hPMS2 defect in HEC-1-A endometrial carcinoma cells. *Cancer Res.* **58:**2978–2981, 1998.
245. Lettieri, T., Marra, G., Aquilina, G., et al. Effect of hMSH6 cDNA expression on the phenotype of mismatch repair-deficient colon cancer cell line HCT15. *Carcinogenesis* **20**, 373–382, 1999.
246. Ahuja, N., Mohan, A. L., Li, Q., et al. Association between CpG island methylation and microsatellite instability in colorectal cancer. *Cancer Res.* **57:**3370–3374, 1997.
247. Lengauer, C., Kinzler, K. W., and Vogelstein, B. DNA methylation and genetic instability in colorectal cancer cells. *Proc. Natl. Acad. Sci. USA* **94:**2545–2550, 1997.
248. Cunningham, J. M., Christensen, E. R., Tester, D. J., et al. Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability. *Cancer Res.* **58:**3455–3460, 1998.
249. Herman, J. G., Umar, A., Polyak, K., et al. Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc. Natl. Acad. Sci. USA* **95:**6870–6875, 1995.
250. Kane, M. F., Loda, M., Gaida, G. M., et al. Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. *Cancer Res.* **57:**808–811, 1997.
251. Bevilacqua, R. A. and Simpson, A. J. Methylation of the hMLH1 promoter but no hMLH1 mutations in sporadic gastric carcinomas with high-level microsatellite instability. *Int. J. Cancer* **87:**200–203, 2000.
252. Endoh, Y., Tamura, G., Ajioka, Y., Watanabe, H., and Motoyama, T. Frequent hypermethylation of the hMLH1 gene promoter in differentiated-type tumors of the stomach with the gastric foveolar phenotype. *Am. J. Pathol.* **157:**717–722, 2000.
253. Simpkins, S. B., Bocker, T., Swisher, E. M., et al. MLH1 promoter methylation and gene silencing is the primary cause of microsatellite instability in sporadic endometrial cancers. *Hum. Mol. Genet.* **8:**661–666, 1999.
254. Glaab, W. E., Hill, R. B., and Skopek, T. R. Suppression of spontaneous and hydrogen peroxide-induced mutagenesis by the antioxidant ascorbate in mismatch repair-deficient human colon cancer cells. *Carcinogenesis* **22:**1709–1713, 2001.
255. Ekbom, A., Helmick, C., Zack, M., and Adami, H. O. Ulcerative colitis and colorectal cancer. A population-based study. *N. Engl. J. Med.* **323:**1228–1233, 1990.
256. Willenbucher, R. F., Aust, D. E., Chang, C. G., et al. Genomic instability is an early event during the progression pathway of ulcerative-colitis-related neoplasia. *Am. J. Pathol.* **154:**1825–1830, 1999.
257. Loeb, K. R. and Loeb, L. A. Genetic instability and the mutator phenotype. Studies in ulcerative colitis. *Am. J. Pathol.* **154:**1621–1626, 1999.
258. Fleisher, A. S., Esteller, M., Tamura, G., et al. Hypermethylation of the hMLH1 gene promoter is associated with microsatellite instability in early human gastric neoplasia. *Oncogene* **20:**329–335, 2001.
259. Chang, C. L., Marra, G., Chauhan, D. P., et al. Oxidative stress inactivates the human DNA mismatch repair system. *Am. J. Physiol. Cell Physiol.* **283:**C148–C154, 2002.
260. Brentnall, T. A., Crispin, D. A., Bronner, M. P., et al. Microsatellite instability in nonneoplastic mucosa from patients with chronic ulcerative colitis. *Cancer Res.* **56:**1237–1240, 1996.

261. Iwaya, T., Maesawa, C., Nishizuka, S., et al. Infrequent frameshift mutations of polynucleotide repeats in multiple primary cancers affecting the esophagus and other organs. *Genes Chromosomes Cancer* **23:**317–322, 1998.
262. Suzuki, H., Harpaz, N., Tarmin, L., et al. Microsatellite instability in ulcerative colitis-associated colorectal dysplasias and cancers. *Cancer Res.* **54:**4841–4844, 1994.
263. Perucho, M. Tumors with microsatellite instability: many mutations, targets and paradoxes. *Oncogene* **22:**2223–2225, 2003.
264. Duval, A. and Hamelin, R. Genetic instability in human mismatch repair deficient cancers. *Ann. Genet.* **45:**71–75, 2002.
265. Lin, H. Y., Wang, X. F., Ng-Eaton, E., Weinberg, R. A., and Lodish, H. F. Expression cloning of the TGF-beta type II receptor, a functional transmembrane serine/threonine kinase. *Cell* **68:**775–785, 1992.
266. Markowitz, S., Wang, J., Myeroff, L., et al. Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science* **268:**1336–1338, 1995.
267. Tomita, S., Miyazato, H., Tamai, O., Muto, Y., and Toda, T. Analyses of microsatellite instability and the transforming growth factor-beta receptor type II gene mutation in sporadic human gastrointestinal cancer. *Cancer Genet. Cytogenet.* **115:**23–27, 1999.
268. Leung, S. Y., Chan, T. L., Chung, L. P., et al. Microsatellite instability and mutation of DNA mismatch repair genes in gliomas. *Am. J. Pathol.* **153:**1181–1188, 1998.
269. Massague, J., Blain, S. W., and Lo, R. S. TGFbeta signaling in growth control, cancer, and heritable disorders. *Cell* **103:**295–309, 2000.
270. Wang, J., Sun, L., Myeroff, L., et al. Demonstration that mutation of the type II transforming growth factor beta receptor inactivates its tumor suppressor activity in replication error-positive colon carcinoma cells. *J. Biol. Chem.* **270:**22,044–22,049, 1995.
271. Grady, W. M., Rajput, A., Myeroff, L., et al. Mutation of the type II transforming growth factor-beta receptor is coincident with the transformation of human colon adenomas to malignant carcinomas. *Cancer Res.* **58:**3101–3104, 1998.
272. Woerner, S. M., Benner, A., Sutter, C., et al. Pathogenesis of DNA repair-deficient cancers: a statistical meta-analysis of putative Real Common Target genes. *Oncogene* **22:**2226–2235, 2003.
273. Duval, A. and Hamelin, R. Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability. *Cancer Res.* **62:**2447–2454, 2002.
274. Malkhosyan, S., Rampino, N., Yamamoto, H., and Perucho, M. Frameshift mutator mutations. *Nature* **382:**499–500, 1996.
275. Lee, W. H., Bookstein, R., Hong, F., Young, L. J., Shew, J. Y., and Lee, E. Y. Human retinoblastoma susceptibility gene: cloning, identification, and sequence. *Science* **235:**1394–1399, 1987.
276. Liu, H., Dibling, B., Spike, B., Dirlam, A., and Macleod, K. New roles for the RB tumor suppressor protein. *Curr. Opin. Genet. Dev.* **14:**55–64, 2004.
277. Li, F. P. and Fraumeni, J. F., Jr. Soft-tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome? *Ann. Intern. Med.* **71:**747–752, 1969.
278. Li, F. P., Fraumeni, J. F., Jr., Mulvihill, J. J., et al. A cancer family syndrome in twenty-four kindreds. *Cancer Res.* **48:**5358–5362, 1988.
279. Li, F. P. and Fraumeni, J. F., Jr. Prospective study of a family cancer syndrome. *JAMA* **247:**2692–2694, 1982.
280. Lustbader, E. D., Williams, W. R., Bondy, M. L., Strom, S., and Strong, L. C. Segregation analysis of cancer in families of childhood soft-tissue-sarcoma patients. *Am. J. Hum. Genet.* **51:**344–356, 1992.
281. Malkin, D., Li, F. P., Strong, L. C., et al. Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* **250:**1233–1238, 1990.
282. Srivastava, S., Zou, Z. Q., Pirollo, K., Blattner, W., and Chang, E. H. Germ-line transmission of a mutated p53 gene in a cancer-prone family with Li–Fraumeni syndrome. *Nature* **348:**747–749, 1990.
283. Varley, J. M. Germline TP53 mutations and Li-Fraumeni syndrome. *Hum. Mutat.* **21:**313–320, 2003.
284. Malkin, D. Li–Fraumeni syndrome, in *The Metabolic and Molecular Bases of Inherited Disease,* 8th ed., Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds., McGraw-Hill, New York, 2001, pp. 849–864.
285. Momand, J., Zambetti, G. P., Olson, D. C., George, D., and Levine, A. J. (1992) The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. *Cell* **69:**1237–1245, 1992.
286. Coleman, W. B., and Tsongalis, G. J. Cancer epidemiology: incidence and etiology of human neoplasms, in *The Molecular Basis of Human Cancer,* Coleman, W. B., and Tsongalis, G. J., eds., Humana Pr, Totowa, NJ, 2002, pp. 3–22.
287. Lynch, H. T. and de la Chapelle, A. Hereditary colorectal cancer. *N. Engl. J. Med.* **348:**919–932, 2003.
288. Nagase, H., Miyoshi, Y., Horii, A., et al. Correlation between the location of germ-line mutations in the APC gene and the number of colorectal polyps in familial adenomatous polyposis patients. *Cancer Res.* **52:**4055–4057, 1992.
289. Su, L. K., Johnson, K. A., Smith, K. J., Hill, D. E., Vogelstein, B., and Kinzler, K. W. Association between wild type and mutant APC gene products. *Cancer Res.* **53:**2728–2731, 1993.
290. Boland, C. R. Hereditary nonpolyposis colorectal cancer (HNPCC), in *The Metabolic and Molecular Bases of Inherited Disease* 8th ed., Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds., McGraw-Hill, New York, 2001, pp. 769–784.
291. Peltomaki, P., Aaltonen, L. A., Sistonen, P., et al. Genetic mapping of a locus predisposing to human colorectal cancer. *Science* **260:**810–812, 1993.
292. Parsons, R., Li, G. M., Longley, M., et al. Mismatch repair deficiency in phenotypically normal human cells. *Science* **268:** 738–740, 1995.
293. Peltomaki, P. and de la Chapelle, A. Mutations predisposing to hereditary nonpolyposis colorectal cancer. *Adv. Cancer Res.* **71:**93–119, 1997.
294. Schwartz, R. A. and Torre, D. P. The Muir–Torre syndrome: a 25-year retrospect. *J. Am. Acad. Dermatol.* **33:**90–104, 1995.
295. Honchel, R., Halling, K. C., Schaid, D. J., Pittelkow, M., Thibodeau, and S. N. Microsatellite instability in Muir-Torre syndrome. *Cancer Res.* **54:**1159–1163, 1994.
296. Turcot, J., Despres, J. P., and St Pierre, F. Malignant tumors of the central nervous system associated with familial polyposis of the colon: report of two cases. *Dis. Colon Rectum* **2:**465–468, 1959.
297. Hamilton, S. R., Liu, B., Parsons, R. E., et al. The molecular basis of Turcot's syndrome. *N. Engl. J. Med.* **332:**839–847, 1995.
298. Miki, Y., Swensen, J., Shattuck-Eidens, D., et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266:**66–71, 1994.
299. Wooster, R., Bignell, G., Lancaster, J., et al. Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378:**789–792, 1995.
300. Sobol, H., Birnbaum, D., and Eisinger, F. Evidence for a third breast-cancer susceptibility gene. *Lancet* **344:**1151–1152, 1994.
301. Antoniou, A. C., Pharoah, P. D., McMullan, G., et al. A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *Br. J. Cancer* **86:**76–83, 2002.
302. Thompson, D., Szabo, C. I., Mangion, J., et al. Evaluation of linkage of breast cancer to the putative BRCA3 locus on chromosome 13q21 in 128 multiple case families from the Breast Cancer Linkage Consortium. *Proc. Natl. Acad. Sci. USA* **99:**827–831, 2002.
303. Easton, D. F., Ford, D., and Bishop, D. T. Breast and ovarian cancer incidence in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. *Am. J. Hum. Genet.* **56:**265–271, 1995.
304. Ford, D., Easton, D. F., and Peto, J. Estimates of the gene frequency of BRCA1 and its contribution to breast and ovarian cancer incidence. *Am. J. Hum. Genet.* **57:**1457–1462, 1995.
305. Venkitaraman, A. R. Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell* **108:**171–182, 2002.
306. Welcsh, P. L., and King, M. C. BRCA1 and BRCA2 and the genetics of breast and ovarian cancer. *Hum. Mol. Genet.* **10:**705–713, 2001.
307. Futreal, P. A., Liu, Q., Shattuck-Eidens, D., et al. BRCA1 mutations in primary breast and ovarian carcinomas. *Science* **266:**120–122, 1994.
308. Shattuck-Eidens, D., McClure, M., Simard, J., et al. A collaborative survey of 80 mutations in the BRCA1 breast and ovarian cancer susceptibility gene. Implications for presymptomatic testing and screening. *JAMA* **273:**535–541, 1995.

309. Gayther, S. A., Warren, W., Mazoyer, S., et al. Germline mutations of the BRCA1 gene in breast and ovarian cancer families provide evidence for a genotype–phenotype correlation. *Nat. Genet.* **11:**428–433, 1995.
310. Albino, A. P. Genes involved in melanoma susceptibility and progression. *Curr. Opin. Oncol.* **7:**162–169, 1995.
311. Bale, S. J., Chakravarti, A., and Greene, M. H. Cutaneous malignant melanoma and familial dysplastic nevi: evidence for autosomal dominance and pleiotropy. *Am. J. Hum. Genet*. **38:**188–196, 1986.
312. Cannon-Albright, L. A., Goldgar, D. E., Meyer, L. J., et al. Assignment of a locus for familial melanoma, MLM, to chromosome 9p13-p22. *Science* **258:**1148–1152, 1992.
313. Goldstein, A. M., Dracopoli, N. C., Ho, E. C., et al. Further evidence for a locus for cutaneous malignant melanoma–dysplastic nevus (CMM/DN) on chromosome 1p, and evidence for genetic heterogeneity. *Am. J. Hum. Genet.* **52:**537–550, 1993.
314. Kamb, A., Gruis, N. A., Weaver-Feldhaus, J., et al. A cell cycle regulator potentially involved in genesis of many tumor types. *Science* **264:**436–440, 1994.
315. Hussussian, C. J., Struewing, J. P., Goldstein, A. M., et al. Germline p16 mutations in familial melanoma. *Nat. Genet.* **8:**15–21, 1994.
316. Kamb, A., Shattuck-Eidens, D., Eeles, R., et al. Analysis of the p16 gene (CDKN2) as a candidate for the chromosome 9p melanoma susceptibility locus. *Nat. Genet.* **8:**23–26, 1994.
317. Ohta, M., Nagai, H., Shimizu, M., et al. Rarity of somatic and germline mutations of the cyclin-dependent kinase 4 inhibitor gene, CDK4I, in melanoma. *Cancer Res.* **54:**5269–5272, 1994.
318. Zuo, L., Weger, J., Yang, Q., et al. Germline mutations in the p16INK4a binding domain of CDK4 in familial melanoma. *Nat. Genet.* **12:**97–99, 1996.
319. Bressac-de-Paillerets, B., Avril, M. F., Chompret, A., and Demenais, F. Genetic and environmental factors in cutaneous malignant melanoma. *Biochimie* **84:**67–74, 2002.
320. Tucker, M. A. and Goldstein, A. M. Melanoma etiology: where are we? *Oncogene* **22:**3042–3052, 2003.
321. Gutmann, D. H. and Collins, F. C. Neurofibromatosis 1, in *The Metabolic and Molecular Bases of Inherited Disease,* 8th ed., Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds., McGraw-Hill, New York, 2001, pp. 877–896.
322. Goldgar, D. E., Green, P., Parry, D. M., and Mulvihill, J. J. Multipoint linkage analysis in neurofibromatosis type I: an international collaboration. *Am. J. Hum. Genet.* **44:**6–12, 1989.
323. Cawthon, R. M., Weiss, R., Xu, G. F., et al. A major segment of the neurofibromatosis type 1 gene: cDNA sequence, genomic structure, and point mutations. *Cell* **62:**193–201, 1990.
324. Viskochil, D., Buchberg, A. M., Xu, G., et al. Deletions and a translocation interrupt a cloned gene at the neurofibromatosis type 1 locus. *Cell* **62:**187–192, 1990.
325. Wallace, M. R., Marchuk, D. A., Andersen, L. B., et al. Type 1 neurofibromatosis gene: identification of a large transcript disrupted in three NF1 patients. *Science* **249:**181–186, 1990.
326. Xu, G. F., O'Connell, P., Viskochil, D., et al. The neurofibromatosis type 1 gene encodes a protein related to GAP. *Cell* **62:**599–608, 1990.
327. Kluwe, L., Siebert, R., Gesk, S., Friedrich, R. E., et al. Screening 500 unselected neurofibromatosis 1 patients for deletions of the NF1 gene. *Hum. Mutat.* **23:**111–116, 2004.
328. Upadhyaya, M., Han, S., Consoli, C., et al. Characterization of the somatic mutational spectrum of the neurofibromatosis type 1 (NF1) gene in neurofibromatosis patients with benign and malignant tumors. *Hum. Mutat.* **23:**134–146, 2004
329. Ponder, B. A. J. Multiple endocrine neoplasia type 2, in *The Metabolic and Molecular Bases of Inherited Disease,* 8th ed., Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds., McGraw-Hill, New York, 2001, pp. 931–942.
330. Carlson, K. M., Dou, S., Chi, D., et al. Single missense mutation in the tyrosine kinase catalytic domain of the RET protooncogene is associated with multiple endocrine neoplasia type 2B. *Proc. Natl. Acad. Sci. USA* **91:**1579–1583, 1994.
331. Haber, D. A. Wilms tumor, in *The Metabolic and Molecular Bases of Inherited Disease,* 8th ed., Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D., eds., McGraw-Hill, New York, 2001, pp. 865–876.
332. Call, K. M., Glaser, T., Ito, C. Y., et al. Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms' tumor locus. *Cell* **60:**509–520, 1990.
333. Gessler, M., Poustka, A., Cavenee, W., Neve, R. L., Orkin, S. H., and Bruns, G. A. Homozygous deletion in Wilms tumours of a zinc-finger gene identified by chromosome jumping. *Nature* **343:**774–778, 1990.
334. Royer-Pokora, B., Beier, M., Henzler, M., et al. Twenty-four new cases of WT1 germline mutations and review of the literature: genotype/phenotype correlations for Wilms tumor development. *Am. J. Med. Genet.* **127A:**249–257, 2004.
335. Etzioni, R., Urban, N., Ramsey, S., et al. The case for early detection. *Nat. Rev. Cancer* **3:**243–252, 2003.
336. Levin, B. Colorectal cancer screening. *Cancer* **72:**1056–1060, 1993.
337. Sidransky, D., Tokino, T., Hamilton, S. R., et al. Identification of ras oncogene mutations in the stool of patients with curable colorectal tumors. *Science* **256:**102–105, 1992.
338. Tobi, M., Luo, F. C., and Ronai, Z. Detection of K-ras mutation in colonic effluent samples from patients without evidence of colorectal carcinoma. *J. Natl. Cancer Inst.* **86:**1007–1010, 1994.
339. Deuter, R., and Muller, O. Detection of APC mutations in stool DNA of patients with colorectal cancer by HD-PCR. *Hum. Mutat.* **11:**84–89, 1998.
340. Eguchi, S., Kohara, N., Komuta, K., and Kanematsu, T. Mutations of the p53 gene in the stool of patients with resectable colorectal cancer. *Cancer* **77:**1707–1710, 1996.
341. Minamoto, T., Mai, M., and Ronai, Z. K-ras mutation: early detection in molecular diagnosis and risk assessment of colorectal, pancreas, and lung cancers—a review. *Cancer Detect. Prev.* **24:**1–12, 2000.
342. Dong, S. M., Traverso, G., Johnson, C., et al. Detecting colorectal cancer in stool with the use of multiple genetic targets. *J. Natl. Cancer Inst.* **93:**858–865, 2001.
343. Powell, S. M., Petersen, G. M., Krush, A. J., *et al.* Molecular diagnosis of familial adenomatous polyposis. *N. Engl. J. Med.* **329:**1982–1987, 1993.
344. Nissum, M., Preuss, D., Harig, A., et al. High-throughput genetic screening using matrix-assisted laser desorption/ionization mass spectrometry. *Psychiatr. Genet.* **12:**109–117, 2002.
345. Bonk, T., Humeny, A., Sutter, C., Gebert, J., von Knebel Doeberitz, M., and Becker, C. M. Molecular diagnosis of familial adenomatous polyposis (FAP): genotyping of adenomatous polyposis coli (APC) alleles by MALDI-TOF mass spectrometry. *Clin. Biochem.* **35:**87–92, 2002.
346. Mao, L., Hruban, R. H., Boyle, J. O., Tockman, M., and Sidransky, D. Detection of oncogene mutations in sputum precedes diagnosis of lung cancer. *Cancer Res.* **54:** 1634–1637, 1994.
347. Andriani, F., Conte, D., Mastrangelo, T., et al. Detecting lung cancer in plasma with the use of multiple genetic markers. *Int. J. Cancer* **108:**91–96, 2004
348. Kimura, T., Holland, W. S., Kawaguchi, T., et al. Mutant DNA in plasma of lung cancer patients: potential for monitoring response to therapy. *Ann. NY Acad. Sci.* **1022:** 55–60, 2004.
349. Almoguera, C., Shibata, D., Forrester, K., Martin, J., Arnheim, N., and Perucho, M. Most human carcinomas of the exocrine pancreas contain mutant c-K-ras genes. *Cell* **53:**549–554, 1988.
350. Urban, T., Ricci, S., Grange, J. D., et al. Detection of c-Ki-ras mutation by PCR/RFLP analysis and diagnosis of pancreatic adenocarcinomas. *J. Natl. Cancer Inst.* **85:**2008–2012, 1993.
351. Berthelemy, P., Bouisson, M., Escourrou, J., Vaysse, N., Rumeau, J. L., and Pradayrol, L. Identification of K-ras mutations in pancreatic juice in the early diagnosis of pancreatic cancer. *Ann. Intern. Med.* **123:**188–191, 1995.
352. Caldas, C., Hahn, S. A., Hruban, R. H., Redston, M. S., Yeo, C. J., and Kern, S. E. Detection of K-ras mutations in the stool of patients with pancreatic adenocarcinoma and pancreatic ductal hyperplasia. *Cancer Res.* **54:**3568–3573, 1994.
353. Tada, M., Omata, M., Kawai, S., et al. Detection of ras gene mutations in pancreatic juice and peripheral blood of patients with pancreatic adenocarcinoma. *Cancer Res.* **53:**2472–2474, 1993.

354. Goh, H. S., Yao, J., and Smith, D. R. p53 point mutation and survival in colorectal cancer patients. *Cancer Res.* **55**, 5217–5221, 1995.
355. Jen, J., Kim, H., Piantadosi, S., et al. Allelic loss of chromosome 18q and prognosis in colorectal cancer. *N. Engl. J. Med.* **331:**213–221, 1994.
356. Borresen-Dale, A. L. TP53 and breast cancer. *Hum. Mutat.* **21:**292–300, 2003.
357. Saitoh, S., Cunningham, J., De Vries, E. M., et al. p53 gene mutations in breast cancers in midwestern US women: null as well as missense-type mutations are associated with poor prognosis. *Oncogene* **9:**2869–2875, 1994.
358. Thorlacius, S., Borresen, A. L., and Eyfjord, J. E. Somatic p53 mutations in human breast carcinomas in an Icelandic population: a prognostic factor. *Cancer Res.* **53:**1637–1641, 1993.
359. Kovach, J. S., Hartmann, A., Blaszyk, H., Cunningham, J., Schaid, D., and Sommer, S. S. Mutation detection by highly sensitive methods indicates that p53 gene mutations in breast cancer can have important prognostic value. *Proc. Natl. Acad. Sci. USA* **93:**1093–1096, 1996.
360. Fielding, L. P. and Pettigrew, N. College of American Pathologists Conference XXVI on clinical relevance of prognostic markers in solid tumors. Report of the Colorectal Cancer Working Group. *Arch. Pathol. Lab. Med.* **119:**1115–1121, 1995.
361. Hammond, M. E., Fitzgibbons, P. L., Compton, C. C., et al. College of American Pathologists Conference XXXV: solid tumor prognostic factors—which, how and so what? Summary document and recommendations for implementation. Cancer Committee and Conference Participants. *Arch. Pathol. Lab. Med.* **124:**958–965, 2000.
362. Henson, D. E., Fielding, L. P., Grignon, D. J., et al. College of American Pathologists Conference XXVI on clinical relevance of prognostic markers in solid tumors. Summary. Members of the Cancer Committee. *Arch. Pathol. Lab. Med.* **119:**1109–1112, 1995.
363. Grignon, D. J. and Hammond, E. H. College of American Pathologists Conference XXVI on clinical relevance of prognostic markers in solid tumors. Report of the Prostate Cancer Working Group. *Arch. Pathol. Lab. Med.* **119:**1122–1126, 1995.

31 Application of Molecular Diagnostics to Hereditary Nonpolyposis Colorectal Cancer

Ian M. Frayling, Lisa Happerfield, Christopher Mattocks, Kim Oakhill, and Mark J. Arends

1. OVERVIEW

Hereditary nonpolyposis colorectal cancer is a tumor predisposition syndrome characterised by a propensity to develop, typically, but by no means exclusively, young-onset colorectal and other cancers *(1)*. The condition was first described in 1913 by the US pathologist Warthin in a comprehensive survey of familial cancer *(2)*. He was stimulated to make this study because his seamstress was depressed at the thought of dying prematurely from bowel or womb cancer, as had many of her relatives. She was a member of Family "G" in his original article, which incidentally contains examples of most of the cancer genetic conditions recognized today *(2)*. Family "G" was rediscovered in the 1960s by Lynch, although it was not at first realized that it was one of Warthin's original families *(3,4)*. Lynch later made a distinction between families with only bowel cancer (Lynch syndrome type 1: site-specific colorectal cancer) and families with several types of cancer, including bowel (Lynch syndrome type 2: family cancer syndrome). Given the propensity to bowel cancer, but without the polyposis characteristic of familial adenomatous polyposis (FAP), the all embracing term "hereditary nonpolyposis colorectal cancer" (HNPCC) is now used *(1)*. However, having to explain all this in the clinic makes the idea of going back to calling it Lynch or family cancer syndrome rather attractive.

Clinically, HNPCC confers an increased risk of many tumor types (Table 1), not all of which are malignant cancers *(1,5,6)*. Diagnosis of the condition is problematic on two grounds. The first, is because it gives few easy clues to the clinician, unlike FAP, which typically presents with hundreds or thousands of adenomatous polyps in the large bowel. Admittedly, in a minority of families, a particular collection of clinical features might lead one to be confident in making a clinical diagnosis, but in the majority, the diagnosis must be empirical and uncertain. Second, familial predisposition to colorectal cancer is common, but HNPCC only accounts for a minority of families. In a meta-analysis of studies of familiality in colorectal cancer, Johns and Houlston found that the risk of colorectal cancer conferred by having just one first-degree relative affected at any age was 2.8 *(7)*. Hence, there is a considerable cohort of families with what has come to be termed moderate (increased) risk of colorectal cancer. Depending on where the line is drawn as to what constitutes significant genetic risk, perhaps 5% or even 10% of colorectal cancer cases are the result of familial predisposition; however, the evidence shows that only about 1% of colorectal cancer cases are the result of HNPCC in populations not subject to a founder effect *(8–15)*. Little is known of the genes conferring a moderate risk of bowel cancer *(16)*. Note that having a moderate risk of bowel cancer needs to be distinguished from having a moderate risk of HNPCC.

The importance of the distinction between HNPCC and families with a moderate risk of bowel cancer lies in the differences it makes in the counseling such families receive *(5)*. This includes advice on the risks they face—not just of bowel cancer, the cancer surveillance program they might be offered, and the resources in mutation detection they will be allocated. Bowel cancer surveillance in HNPCC is currently recommended to be colonoscopy every 2 yr from the age of 25. The appropriate regime in moderate-risk families has yet to be determined, but, for example, our current guidelines recommend colonoscopy every 5 yr from age 45, considerably different from that for HNPCC. Also, counseling of moderate-risk families does not need to extend to tumors other than colorectal and it does not need to include the possibility of germline mutation detection and predictive testing. In these circumstances, therefore, any test able to exclude HNPCC is valuable.

The International Collaborative Group on HNPCC (ICG-HNPCC) was set up to study the condition, and in 1990, it defined a set of criteria at its first formal meeting in Amsterdam in order to select families for linkage studies: these were the so-called 'Amsterdam Criteria' (Table 2). It should be noted that

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

Table 1
Types and Lifetime Risks of Tumours in HNPCC

Site	*Risk in HNPCC*	*Approximate increase over general population*	*Lifetime UK General Population Risk*[a]
Colorectum	45% (by 50 yr)	(× ~200)	
Males	70% (by 70 yr)		
	80% (lifetime)	× 16	5.5% (1 in 18)
Females	20% (by 50 yr)	(× ~100)	
	35% (by 70 yr)		
	40% (lifetime)	× 8	5.0% (1 in 20)
Endometrium	10% (by 50 yr)		
	40% (by 70 yr)		
	50% (lifetime)	× 35	1.4% (1 in 70)
Gastric	~5%	× 2~3	Males: 2.3% (1 in 45) Females: 1.2% (1 in 90)
Ovarian	~4%	× 2–3	2.1% (1 in 50)
Pancreas	2%?	× 2–3?	1.0% (1 in 100)
Urothelial (especially ureter/renal pelvis)	2–6%	× 2–3	Males: 3.3% (1 in 30) Females: 1.3% (1 in 80)
Renal (not urothelial)	1–2%	× 2–3	Males: 1.1% (1 in 90) Females: 0.6% (1 in 160)
Brain/CNS	3%	× 4	0.7% (1 in 150)
Lymphoproliferative (not Hodgkin's)	2–3%	× 2–3	1.3% (1 in 75)
Small bowel	1%	× ~300	<0.003%
Hepatobiliary system (gallbladder/bile duct)	1%?	× 2–3?	

[a]UK National Office of Population and Census Studies (OPCS) data, 1996.
Source: ref. *5*.

Table 2
The Amsterdam Criteria 2 (AC2) for Selection of Possible HNPCC Families

1. At least three separate relatives with colorectal cancer (CRC) or HNPCC-related cancer, including endometrial (HRC: see criterion 1 and 2 in Table 3)[a]
2. One relative must be a FDR of the other two[b]
3. At least two successive generations affected
4. At least one cancer (CRC or HRC) diagnosed <50 yr
5. FAP excluded in CRC case(s)
6. Tumors pathologically verified

[a]The first version of the criteria (AC1) specified CRC alone (i.e., it did not take into account endometrial and other HRC).
[b]Affected individuals must all be on the same side of the family (e.g., maternal or paternal relatives of the consultand).
Source: ref. *15*.

these criteria were *not* to enable clinical diagnosis of the condition, but solely for research purposes. The ICG-HNPCC has subsequently published a set of guidelines for clinical diagnosis (Table 3) *(15)*. It should be noted that the characteristic feature of HNPCC is a germline mutation in one of the DNA mismatch repair (MMR) genes and that the tumors have microsatellite instability (MSI) and can show abnormality of MMR gene expression evident on immunohistochemistry (IHC) *(1,17–28)*. Therefore, we have taken a combined multidisciplinary approach to the diagnosis of HNPCC *(29–36)*.

Table 3
ICG-HNPCC Criteria for Clinical Diagnosis of HNPCC

1. Familial clustering of colorectal and/or endometrial cancer[a]
2. Associated HNPCC-related cancers (HRC): gastric, ovary, ureter/renal pelvis, brain, small bowel, hepatobiliary tract and skin (sebaceous tumors)[a]
3. Development of cancer at an early age.
4. Development of multiple cancers.
5. Features of colorectal cancer:
   (a) predilection for proximal (right-sided) colon;
   (b) improved survival;
   (c) multiple colorectal cancers;
   (d) increased proportion of mucinous tumors, and tumors with marked host–lymphocytic infiltration and lymphoid aggregation at the tumor margin
6. Features of colorectal adenoma:
   (a) the numbers vary from one to a few;
   (b) increased proportion of adenomas with villous histology and
   (c) high-grade dysplasia,
   (d) probably rapid progression from adenoma to carcinoma.
7. High frequency of microsatellite instability (MSI)
8. Immunohistochemistry: loss of MLH1, MSH2, or MSH6 protein expression in tumors.
9. Germline mutation in MMR gene: *MSH2, MLH1, MSH6, PMS2*.
   Any combination of 1–9 might be present, but 9 alone defines HNPCC.

[a]Endometrial cancer is classed as an HRC
Source: ref. *15*.

2. DNA MISMATCH REPAIR

There is a host of DNA repair systems available to organisms, each one directed at a particular class of DNA lesions but with considerable overlap between them, so that the organism is not wholly dependent on any one. DNA damage occurs when agents that impinge on it cause an alteration in its structure. Often, this amounts to changes in the bases by breaking and making of covalent bonds, caused by (e.g., ultraviolet [UV] light, oxidation, or alkylating agents). Such damaged bases are able to make abnormal pairings. For example, 8-hydroxyguanine (8oxo-dG), a product of oxidative damage, is equally able to pair with A as it is with C and is thus highly mutagenic. Sometimes, however, a more subtle form of damage can occur in which, for one reason or another, normal bases (A, C, G, or T) become abnormally paired together, causing a point mismatch (e.g., A:G rather than A:T. Where nucleotides are gained or lost in one strand relative to the other, the consequent unmatched stretch of bases forms an insertion–deletion loop (IDL), which constitutes a second form of mismatch. Both point mismatches and IDL are detected and made good by the DNA MMR system *(26,27)*.

In *Escherichia coli,* the MMR system consists of three proteins: mutH, mutS, and mutL (an animated cartoon illustrating the process can be seen at http://www.jci.tju.edu/rfishel/research.html). MutH is a strand discrimination endonuclease that senses the dam-methylation status of the bacterium's DNA, and, thus, it has no equivalent in higher organisms. In consequence, their MMR system is composed of homologues of mutS and mutL; thus, MSH2 is human mutS homolog 2, and MLH1 is human mutL homolog 1. Some homologs of mutL were identified in yeast as mutants causing abnormalities of postmeiotic segregation of chromosomes; hence, PMS2 is the human homolog of yeast postmeiotic segregation 2.

Point mismatches and single-nucleotide IDL are bound by a complex consisting of MSH2 and MSH6 (hMutSalpha), whereas a complex of MSH2 and MSH3 (hMutSbeta) binds to IDLs of two to eight nucleotides, but it should be noted that there is overlap in the specificities of these two complexes and hence redundancy in activity *(37–40)*. Either hMutSalpha or hMutSbeta is then able to recruit either hMutLalpha (MLH1 and PMS2) or hMutLbeta (MLH1 and PMS1) *(41,42)*. Resolution of the mismatched nucleotide or IDL is then achieved. As MSH2 and MLH1 are the common components, then loss of either will abrogate all MMRs, whereas loss of one of the other components (e.g., MSH6) will only diminish the activity of MMR against point/single-nucleotide IDLs. This is probably why most HNPCC families are found to have a mutation in either *MSH2* or *MLH1*, whereas *MSH6* families are much less common and have a somewhat different propensity to colorectal vs endometrial cancer *(43–46)*. Recent work has shown that the N-terminus of MLH1 is responsible for the binding of hMutSalpha to hMutLalpha, whereas the C-terminus of MLH1 interacts with PMS2 *(47)*.

2.1. IMMUNOHISTOCHEMICAL ANALYSIS OF MMR PROTEIN EXPRESSION IN TISSUE SECTIONS Immunohistochemistry (IHC) is a technique that allows the specific staining within tissue sections of particular proteins. A specific, primary antibody, either monoclonal or polyclonal, binds to the target protein in a tissue section and is detected by means of a secondary antibody (e.g., rabbit anti-mouse, if the primary antibody were a mouse-derived monoclonal), covalently bound to which is a visualization system, usually an "ABC" system with biotin bound to the secondary antibody, followed by addition of an avidin–biotin complex (ABC) containing biotinylated peroxidase enzyme (sometimes the peroxidase enzyme or a fluorescent dye might be bound to the secondary antibody). The peroxidase then converts a substrate into an insoluble colored product (upon oxidation by the peroxidase enzyme with buffered hydrogen peroxide, DAB [3,3-diaminobenzidine] forms a very stable, brown endproduct) at the site of the target antigen. Thus, IHC is able to demonstrate not just the presence of a protein in a tissue but also its histological distribution at the subcellular level (e.g., nuclear, cytoplasmic, membranous, extracellular matrix).

In HNPCC tumors that have lost MMR function, IHC can be used to show abnormal MMR protein expression *(24,36,48)*. HNPCC sufferers inherit one germline mutant MMR allele and one normal, wild-type MMR allele. During tumor formation, the normal allele is inactivated by mutation or loss, or silenced by promoter hypermethylation, thus leaving no expression of functional alleles. Thus, MMR abnormal expression in HNPCC tumors could be detected by IHC in two patterns: either complete loss of expression (when there is no expression of that MMR protein or only expression of a truncated protein to which the antibody does not bind) or patchy, weak expression (if the mutation generates a protein with a shorter half-life, or a prematurely truncated but variably stable protein, or a protein with alterations to the epitope recognized by the antibody [e.g., missense mutation]). This patchy weak staining might often be seen in the cytoplasm rather than the nucleus, whereas the adjacent normal epithelium or intratumoral activated lymphocytes show the usual pattern of strong nuclear immunopositivity for the MMR protein (because of the expression of the wild-type allele). This suggests that destabilization of the MMR protein complexes has occurred and the proteins are no longer bound to the nuclear DNA *(49)*. It is important to realize, however, that about one-third of HNPCC-related tumors do not exhibit any abnormality on analysis by IHC even though they have lost MMR function, as manifested by MSI, and this might be the result of mutations that functionally inactivate the MMR protein but allow its expression as a stable protein with nuclear localization and intact epitope.

2.1.1. WHICH MMR PROTEINS TO TEST? As a minimum, IHC for MSH2 and MLH1 proteins should be performed, but we have taken the approach of including IHC for MSH6 and PMS2 proteins *(24,44,48–51)*. Abnormality of MSH6 frequently accompanies abnormality of MSH2, as MSH6 is secondarily destabilized by a lack of MSH2 with which to form hMutSalpha complexes *(40,42)*. Similarly, abnormality of PMS2 frequently accompanies abnormality of MLH1, as PMS2 is secondarily destabilized by a lack of MLH1 with which to form hMutLalpha complexes *(26,47)*. Thus, in families with MSH2 mutations, the pattern of IHC across a number of tumors could be no abnormality, or abnormality of MSH2 or MSH2 plus MSH6, but not MSH6 alone. However, tumors

from a family with an MSH6 mutation can be expected to show consistent loss of MSH6 alone, although they can sometimes show loss of MSH2 and MSH6 *(45)*.

The IHC protocols (including antigen retrieval procedures) for MSH2 and MLH1 are essentially those determined to perform best in multicenter trials carried out under the auspices of the ICG-HNPCC with local minor alterations for automation of immunostaining *(52,53)*. Similarly, the IHC protocol (including antigen retrieval procedures) we use for MSH6 is essentially the one used in a recent ICG-HNPCC trial (data not yet reported) with minor local alterations for automation of immunostaining *(49)*. The protocol (including antigen retrieval procedures) for PMS2 is one we have determined in our own laboratory, optimized for automation of immunostaining *(49)*. Automation of immunostaining (e.g., using the DAKO TechMate 500) allows highly consistent and reproducible immunostaining procedures, permitting comparison of the immunostaining patterns between different tissues and cases. This is particularly useful when suboptimally fixed tumor tissues must be analyzed, as the four anti-MMR antibodies are particularly fixation sensitive and care must be taken to evaluate immunostaining patterns only in well-fixed regions of the tissue section (e.g., where the adjacent normal epithelium or the intratumoral or peritumoral lymphoid cells show clear, strong nuclear immunopositivity for all of the MMR proteins).

2.1.2. IMMUNOHISTOCHEMISTRY: PRACTICAL POINTS We would recommend that all IHC be carried out using some form of automated staining machine (e.g. Dako TechMate 500 or similar). Manual IHC is prone to significant variation both within and between batches. Automated immunostaining generates highly consistent and reproducible patterns both within and between different batches. The antigenic epitopes for the four anti-MMR antibodies are particularly fixation sensitive and immunostaining patterns should only be assessed in well-fixed regions of the tissue section (e.g., where the adjacent normal epithelium or the intra-tumoral or peritumoral lymphoid cells show clear, strong nuclear immunopositivity for all four of the MMR proteins).

We report MMR IHC findings in tumors as either of the following:

1. Normal: showing strong immunopositivity of the tumor cell nuclei for all four MMR proteins, similar in staining intensity to that of the adjacent normal epithelium or intratumoral lymphoid cells (Fig. 1A,B).
2. Negative: showing complete loss of staining, relative to the strongly intensive immunopositivity of the adjacent normal epithelium or intratumoral lymphoid cells (Fig. 1A,B).
3. Patchy/weak: showing abnormal, weak, heterogeneous immunostaining intensity, relative to the strongly intensive immunopositivity of the adjacent normal epithelium or intratumoral lymphoid cells, and this might appear cytoplasmic rather than nuclear in location. We have found that this weak/patchy pattern of immunostaining runs in a minority of families, in that different tumors from different affected members of the same family tend to show a very similar heterogeneous immunostaining pattern (Fig. 1C).

If type 1, then our reports conclude with a statement that evidence of abnormal MMR expression has not been found, and the covering statement is made that MSI studies should be carried out to include or exclude MMR abnormality in the tumor. Otherwise, if type 2 or 3, our reports conclude with a statement that evidence of abnormal MMR expression has been found and that this is compatible with, but not necessarily diagnostic of, HNPCC. The comment is also made that MSI studies are not needed.

We have found that a small minority of tumors, about 1%, show unusual combinations of abnormal MMR protein expression (e.g., loss of both MSH2 and MLH1, or three out of the four proteins). It is suggested that the IHC be repeated and then a clinico-pathological review of the case be carried out, as necessary, to decide whether to proceed with testing other tumors from the same individual or family or to proceed directly with mutation detection.

2.2. MICROSATELLITE INSTABILITY At an early stage in the discovery of the molecular basis of HNPCC, linkage was established to chromosome 2p. As HNPCC families showed a dominant pattern of inheritance, it was reasonable to suppose that the tumors from families linked to 2p might show loss of heterozygosity (LOH) of markers on 2p. When this was done, however, it was not a loss but a paradoxical gain of alleles that was observed, originally ascribed to replication errors (RER) *(21–23)*. This clue was picked up by yeast biochemists who recognized that this was a manifestation of loss of MMR, thus leading to the identification of *MSH2* on 2p and *MLH1* on 3p *(18–20)*. Originally termed RER, this phenomenon is now called microsatellite instability (MSI); however, it should be noted that whereas mutations in cells that have lost MMR are more likely in repetitive DNA, they are, in fact, occurring all over the genome *(54)*.

The first microsatellites were dinucleotide (CA_n) repeats, but it was later found that MSI could be observed at mononucleotide, dinucleotide, trinucleotide, and tetranucleotide repeats. Individual microsatellites vary in their propensity to exhibit instability upon loss of MMR, and this variation is also dependent on the tumor type being studied *(55–64)*.

2.2.1. CATEGORIES OF MSI Early studies seemed to show that if a number of microsatellite markers were tested, tumors varied in the proportion that showed MSI *(21–23)*. There were tumors that showed MSI in a large proportion of markers and these were classed as having MSI-H (for high-frequency MSI). Then, there were tumors that showed no instability, termed MSS (for microsatellite stable), and, finally, a third class that showed MSI in a small proportion of markers, termed MSI-L (for low-frequency MSI). In terms of their biological properties, tumors with MSI-L appeared to be more like MSS tumors than MSI-H tumors. At a meeting organized at the NCI (Bethesda, MD USA), guidelines were laid down as to the proportion of markers tested needing to show instability before a diagnosis of MSI could be made in that tumor (Table 4) *(65)*. Given the ambiguity in the minimum proportion of markers needing to exhibit instability (30–40%), it is useful to calculate the proportion of markers (Table 5) and the absolute number of markers needed to make a diagnosis of MSI (Table 6), given any total number of markers tested. The cutoff figure of the proportion of markers needed to distinguish MSI, if testing between 5 and 12 markers, comes out at 29%, which is useful

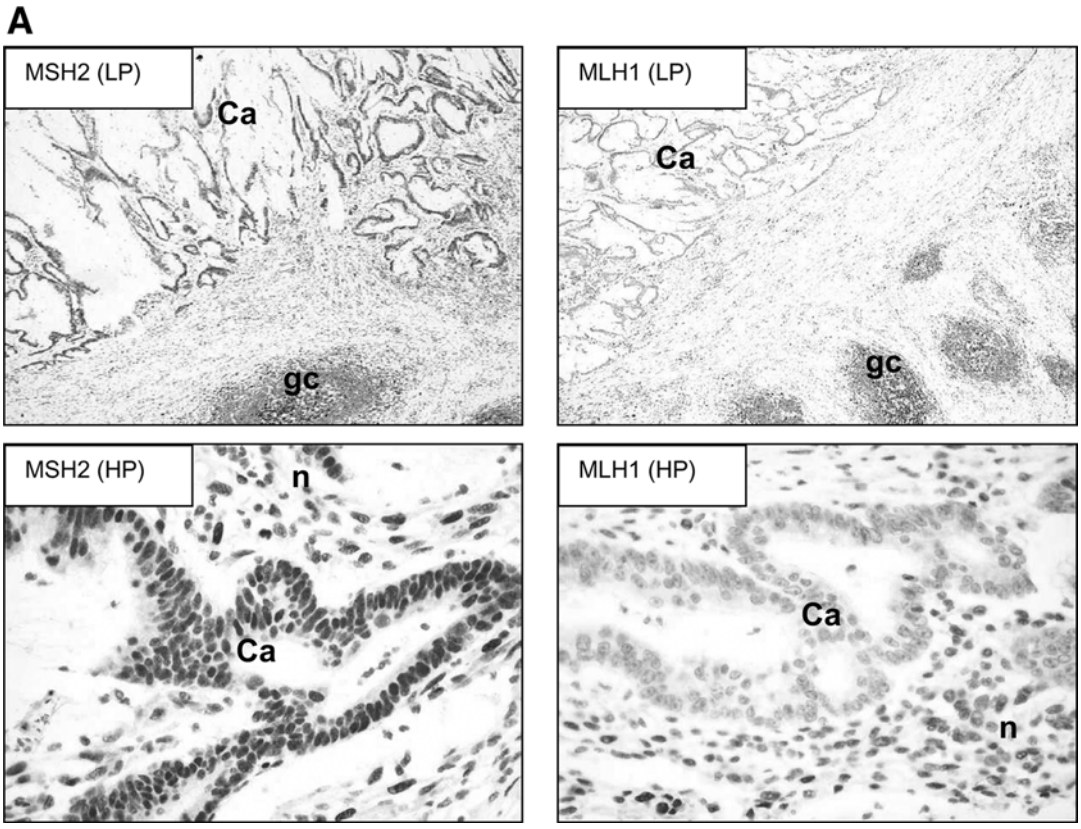


Fig. 1. Examples of DNA mismatch repair immunohistochemistry. **(A)** Loss of MLH1 in a mucinous colon cancer. Panels show low-power (LP) and high-power magnification (HP) of IHC of MSH2 and MLH1. Normal tissue germinal centers (gc) expresse both proteins, as do normal lymphocytes (n) in the stroma surrounding the cancer (Ca), and thus acts as the ideal internal control. The cancer, however, has clearly lost expression of MLH1 while retaining MSH2. **(B)** Loss of MSH2 in both colon and endometrial cancers from the same patient. Panels show IHC of MSH2 and MLH1 from a colon cancer and endometrial cancer from the same patient. Both tumors show concordant loss of MSH2. Note how the normal glandular crypts in both the colonic mucosa and endometrium stain positively for both MLH1 and MSH2 (n). **(C)** Patchy/weak staining is seen in some tumors from some families. Top panels show IHC of MSH2 and MLH1 in an endometrial cancer from a woman in a large HNPCC family. MSH2 is expressed in normal endometrial crypts (n) and the cancer (Ca), but although MLH1 is expressed in the normal crypts, its expression in the cancer is patchy and weak, but not completely absent. The bottom left Panel is a young-onset transitional cell papilloma (Pa) from a second member of the family (18 yr), which, like the female's endometrial cancer, also shows patchy/weak expression of MLH1. A cecal cancer from a third member of the family also shows patchy/weak expression of MLH1 (not shown). The bottom right Panel, however, shows a transitional cell carcinoma (TCCa) from a fourth member of the family, which is the only one to have completely lost expression of MLH1 (Ca), although normal stromal and endothelial cells retain expression (n). All four of these individuals have the same germline mutation: *MLH1* IVS8 + 3G > A (677 + 3G > A). (Figure appears in color in insert following p. 172.)

if data are to be analyzed on a computerized spreadsheet or database.

There has been an ongoing debate as to whether MSI-L as a distinct entity exists. This was significantly addressed recently by two groups performing essentially the same experiment: testing a large set of tumors using a large set of microsatellites. When this was done, it was found that if enough markers are tested, eventually one is found to be unstable, and, hence, all tumors exhibit at least MSI-L *(66,67)*. Moreover, those tumors with MSI-H were found to be biologically distinct from those with MSI-L, but no convincing evidence could be found to justify a separate classification of MSI-L. These findings make a considerable difference in the clinical diagnostic reporting of MSI in tumors, as it is no longer necessary to attempt to interpret MSI-L and reports can simply state that the tumor either does have MSI or does not have MSI. Therefore, we report MSI status in tumors according to Table 7, which gives the absolute number of markers with instability needed to diagnose MSI, given that the distinction between MSI-L and MSS is not useful in the diagnosis of HNPCC (although it might still prove to be so in the diagnosis of other forms of hereditary bowel cancer).

However, there are two caveats. First, the occasional tumor is found that only shows MSI-L but does have abnormality of MMR protein expression on IHC. Given the probabalistic nature of whether a marker is affected by instability or not, it is to be expected that some tumors with MSI will be observed in the lower end of the distribution. In such cases, the interpretation must take into account the familial setting of the individual and, for example, the type or age of onset of the tumor. Second, individual markers vary in their sensitivity to instability, particularly in tumors other than colorectal, and needs to be taken into account. In addition, instability at mononucleotide repeats is more significant than that at dinucleotide repeats *(58)*. In our experience, there is a small but definite proportion of tumors

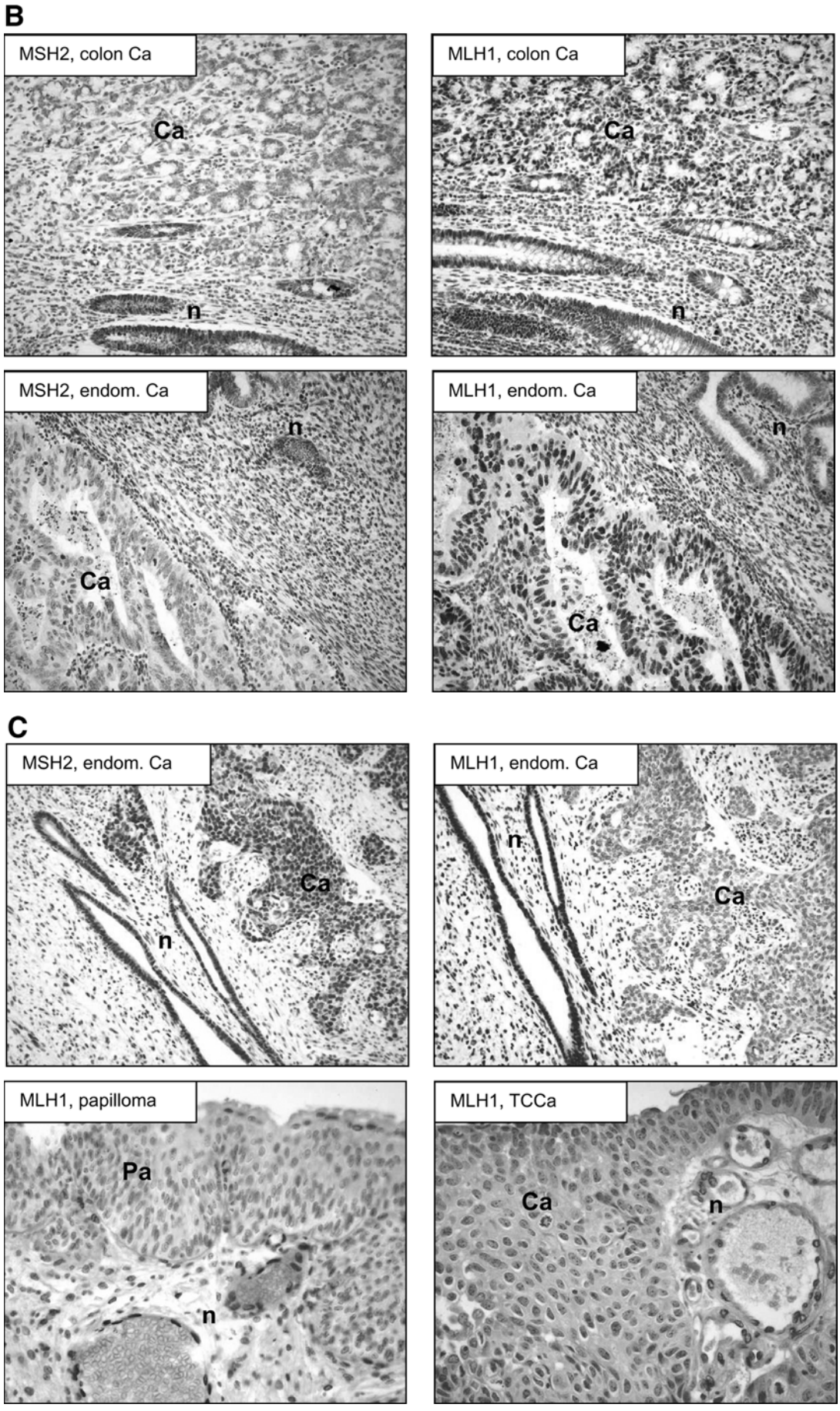


Fig. 1. *(Continued)*

Table 4
NCI Microsatellite Instability Diagnosis Guidelines

Total no. of markers tested	*No. /proportion of markers tested showing MSI*	*MSI category*
Five loci analyzed	2+	MSI-H
	1	MSI-L
	0	MSS
More than five loci analyzed	30–40%+	MSI-H
	<30–40%	MSI-L
	0	MSS

Source: ref. *65.*

Table 5
Proportion of Markers with Instability Needed to Make a Diagnosis of MSI, Based on NCI 1998 Guidelines

	5	***6***	***7***	***8***	***9***	***10***	***11***	***12***
40%	2.0	2.4	2.8	3.2	3.6	4.0	4.4	4.8
30%	1.5	1.8	2.1	2.4	2.7	3.0	3.3	3.6
MSI-H >29% cutoff	40%	33%	43%	38%	33%	30%	36%	33%
MSI-L	20%	17%	29%	25%	22%	20%	27%	25%

Source: ref. *65.*

Table 6
Absolute Number of Markers with Instability Needed to Make a Diagnosis of MSI, Based on NCI 1998 Guidelines

	5	***6***	***7***	***8***	***9***	***10***	***11***	***12***
MSI-H	≥2	≥2	≥3	≥3	≥3	≥3	≥4	≥4
MSI-L	1	1	1–2	1–2	1–2	1–2	1–3	1–3
MSS	0	0	0	0	0	0	0	0

Source: ref. *66.*

Table 7
Absolute Number of Markers with Instability Needed to Make a Diagnosis of MSI, Based on NCI 1998 Guidelines and Removing Distinction Between MSI-L and MSS

	5	***6***	***7***	***8***	***9***	***10***	***11***	***12***
MSI	≥2	≥2	≥3	≥3	≥3	≥3	≥4	≥4
Not MSI	0–1	0–1	0–2	0–2	0–2	0–2	0–3	0–3

Source: refs. *65–67.*

with abnormal IHC (often correlating with other tumors from the same individual and/or family) but that have instability at <30% of markers. However, of the markers that are affected, at least one is a mononucleotide repeat. Thus, we class tumors that have instability at one or more mononucleotide repeats, but less than 30% of all markers tested, as having MSI. If uncertainty exists and other tumors from the same individual or family can be obtained, it is well worth testing them, as it will usually be found that a convincing pattern then emerges.

2.2.2. HOW MANY MICROSATELLITES TO TEST? Initially, different groups used different numbers of microsatellites to test for MSI. In an attempt to facilitate comparison between studies, the ICG-HNPCC organized a meeting at the National Cancer Institute (Bethesda, MD). Here, it was decided to recommend a 'core' set of five microsatellites to be used in all studies, with the choice of selecting others from a secondary set of 19 markers (Tables 5–7) *(65)*. It is often confusing to those new to the subject that although an individual microsatellite can exhibit instability, the diagnosis of MSI in an individual tumor is dependent on more than a certain proportion of all the markers tested showing instability. It is recommended that not less than five markers be tested to assess MSI *(56,68,69)*. Testing more markers, especially if the polymerase chain reactions (PCRs) can be multiplexed and analyzed using a fluorescent DNA analyzer, allows for a proportion of microsatellites to fail to amplify but also gives considerable confidence once a pattern starts to appear across a number of markers. Also, given that most markers have been decided upon on the basis of their efficiency at detecting MSI in colorectal tumors and hence are less efficient in noncolorectal tumors, testing more than five compensates for this.

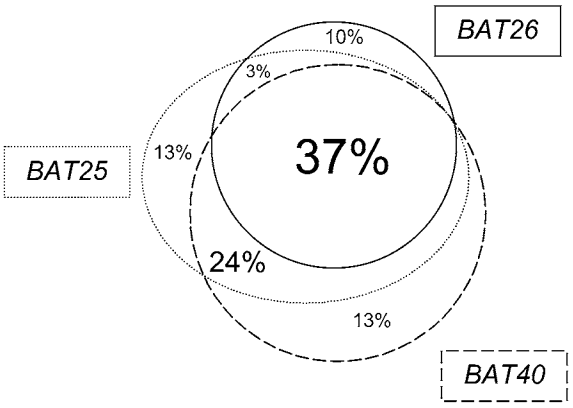


Fig. 2. No single mononucleotide marker detects MSI in all tumors from possible HNPCC families. Proportion of tumors with MSI showing instability at different mononucleotide markers ($n = 30$).

2.2.3. WHICH MICROSATELLITES TO TEST? Early studies used a variety of different microsatellites to test for MSI. In an attempt at standardization, the Bethesda recommendations were for a core panel of five markers, to include two mononucleotide repeats (*BAT25* and *BAT26*) and three dinucleotide repeats (*D5S346*, *D2S123,* and *D17S250*), plus 19 other microsatellites from which to choose additional markers *(65)*. Others have suggested that testing either *BAT25* or *BAT26* alone would be sufficient *(63)*. This might be adequate for the diagnosis of MSI in sporadic tumors, in which the mechanism of MMR loss is distinct from that in HNPCC, in that it is largely the result of methylation of the *MLH1* promoter (see subsection 2.2.4.). However, we have found that *BAT25* or *BAT26* alone would be insufficient in diagnosing familial cases, possibly the result of HNPCC (Fig. 2). Bearing in mind all of the various factors, we have settled on a panel of 10 microsatellites, 3 mononucleotide repeats, and 7 dinucleotide repeats (Table 8) *(61,70–73)*. Five of these can be performed in a single multiplex PCR, reducing the number of separate PCRs to six *(71)*. By analyzing them on a fluorescent DNA analyzer (ABI 377 or 3100) and labeling amplicons of similar sizes with

Table 8
Microsatellite Repeats for MSI Testing

Microsatellite	*Repeat*	*Location*	*Refs.*
BAT25	$(A)_{25}$	4q12, within intron of c-kit	*61*, *70*
BAT40	$(A)_{40}$	1p13, within intron of *HSD3B1*	*70*
BAT26	$(A)_{26}$	2p16, within intron 5 of *MSH2*	*61*, *70*, *71*
ACTC	$(AC)_{24}$	15q11-q14	*71*
D5S406	$(CA)_{27}$	5p15	*71*
D5S107[a]	$(CA)_9AA(CA)_{19}(GA)_7$	5q11-q13	*71*
D13S153	$(CA)_{25}$	13q14	*71*
D2S123	$(CA)_{13}TA(CA)_{15}$	2p16, near *MSH2* and *MSH6*	*61*
D5S346[b]	$(CA)_{15}$	5q22, within 35 kb of *APC*	*72*
D17S250[c]	$(AC)_{23}$	17q12	*73*

[a]also known as *Mfd27*.
[b]also known as *Mfd15*.
[c]also known as *LNS-CA*.

different fluorophores, it is possible to run all 10 microsatellites in a single lane/capillary if very high throughput is required.

2.2.4. SIGNIFICANCE OF MSI It is not just tumors arising in HNPCC patients that can lose MMR. Approximately 15–20% of most common cancers, usually those of the same type as found in HNPCC but sporadic in origin, are found to have lost MMR *(23,27,74)*. Thus, a good proportion (approx 15–20%) of colon, endometrial, and ovarian tumors, *inter alia*, exhibit MSI. In this context, colon cancer does not include rectal cancers, the biology of which is distinct (*see* below). The mechanism of loss of MMR in sporadic tumors, at least colon cancers, is generally different from that seen in HNPCC patients. In sporadic colon and endometrial cancers, loss of MMR typically occurs by hypermethylation of the *MLH1* promoter, downregulating its expression *(75–77)*. It used to be thought this was a mechanism exclusive to sporadic tumors, but a proportion of HNPCC colon cancers also lose MLH1 via methylation, so this is not quite the discriminant it was hoped it might be *(78–80)*. There are a number of ways in which *MLH1* promoter methylation can be tested for, but it must be borne in mind that they measure somewhat different events *(81–83)*. A test that depends on cleavage at a particular CpG site between two primer binding sites will report methylation even though none of the other CpG sites in the region in that molecule are methylated. This would also be the case for, for example, an ARMS test, dependent on binding of a methylation-specific primer to a particular CpG, after sulfite modification. These assays might report methylation in, say, 10% of the DNA tested, but the actual methylated CpG sites might be randomly distributed, with no DNA molecules actually being completely or largely methylated. This has been addressed by the design of a real-time PCR assay (MethyLight), in which all the primers involved (forward, reverse, and reporter) bind to sites containing at least three CpGs *(83)*. Thus, the MethyLight assay (Chapter 7) only reports methylation when all of the CpG sites on a particular DNA molecule are methylated, a situation that is likely to be more biologically relevant *(81–83)*.

It should be noted that rectal cancers very rarely, if at all, lose MMR sporadically, and, thus, a rectal cancer with MSI or abnormal MMR IHC is excellent evidence of HNPCC *(84)*. Similarly, colorectal adenomas rarely show MSI outside of HNPCC, so the finding of an adenoma with MSI confers a high predictive value for HNPCC *(55,74,85,86)*.

2.2.5. MICROSATELLITE INSTABILITY: PRACTICAL POINTS The average length of DNA isolated from fixed tissues is often quoted as to be up to 650 bp, depending on, for example, the type and age of the sample and the type of fixative used. In our experience, the maximum amplicon size for DNA extracted from wax-embedded tissue is about 250–270 bp. Those histological fixatives that cause crosslinking, such as osmic acid or mercuric salts, are known causes of subsequent PCR failure. At first look, not all unstable microsatellites might appear so. It helps considerably to compare the traces from a tumor with normal tissue and to do this comparison with data from a fluorescent DNA analyzer using appropriate software.

As experience is gained, it will become apparent that stability is obvious and deviations from normal are evidence of instability (Fig. 3). We have yet to find a tumor with abnormal IHC that does not exhibit MSI; therefore, carry out initial pilot experiments using DNA extracted from such blocks and build up a bank of material from such tumors for control purposes. It might be thought that cell lines with MSI will be good controls, but, in fact, they are difficult to interpret because of the lack of corresponding normal DNA. One of the advantages of testing 10 loci is that if 1 locus is doubtful, then it can be scored as uninformative for the purposes of MSI detection. It is often the case that although there might be initial uncertainty about how to score a particular locus, this is resolved upon finding that most of the other loci tested in the same tumor show clear-cut instability. It is not uncommon to observe subclonal evolution when samples are taken from two or more different regions of the same tumor (Fig. 3C).

Be aware that although *BAT26* is pseudomonomorphic in Caucasian populations (i.e., almost all individuals have an allele amplifying with a modal amplicon size of 119 bp using the primers stated) in other populations this locus is polymorphic and, thus, caution is required if corresponding normal DNA is not available *(87–89)*. The other two mononucleotide repeats, *BAT25* and *BAT40*, are polymorphic and, thus, heterozygous in some individuals. This can be confused with instability, but it will be found that in cases of instability, the normal smooth symmetrical outline of the stutter bands around the modal amplicon is lost, whereas stable alleles retain this pattern.

The predictive value of MSI in terms of HNPCC depends on a number of factors. Finding MSI in a solitary colon cancer, of average age of onset (i.e., 70's) in an individual without a family history or personal history of other HNPCC-related tumors almost certainly indicates a sporadic tumor. Whereas finding MSI in a young-onset HNPCC-related tumor type from an individual with a family history consistent with HNPCC is significant. In any circumstances, finding MSI in a tumor type that does not show a high rate of sporadic MSI (e.g., a colorectal adenoma or rectal carcinoma) is significant *(74,84,86)*. Finding MSI in two or more tumors from the same individual or family also increases the significance *(74)*.

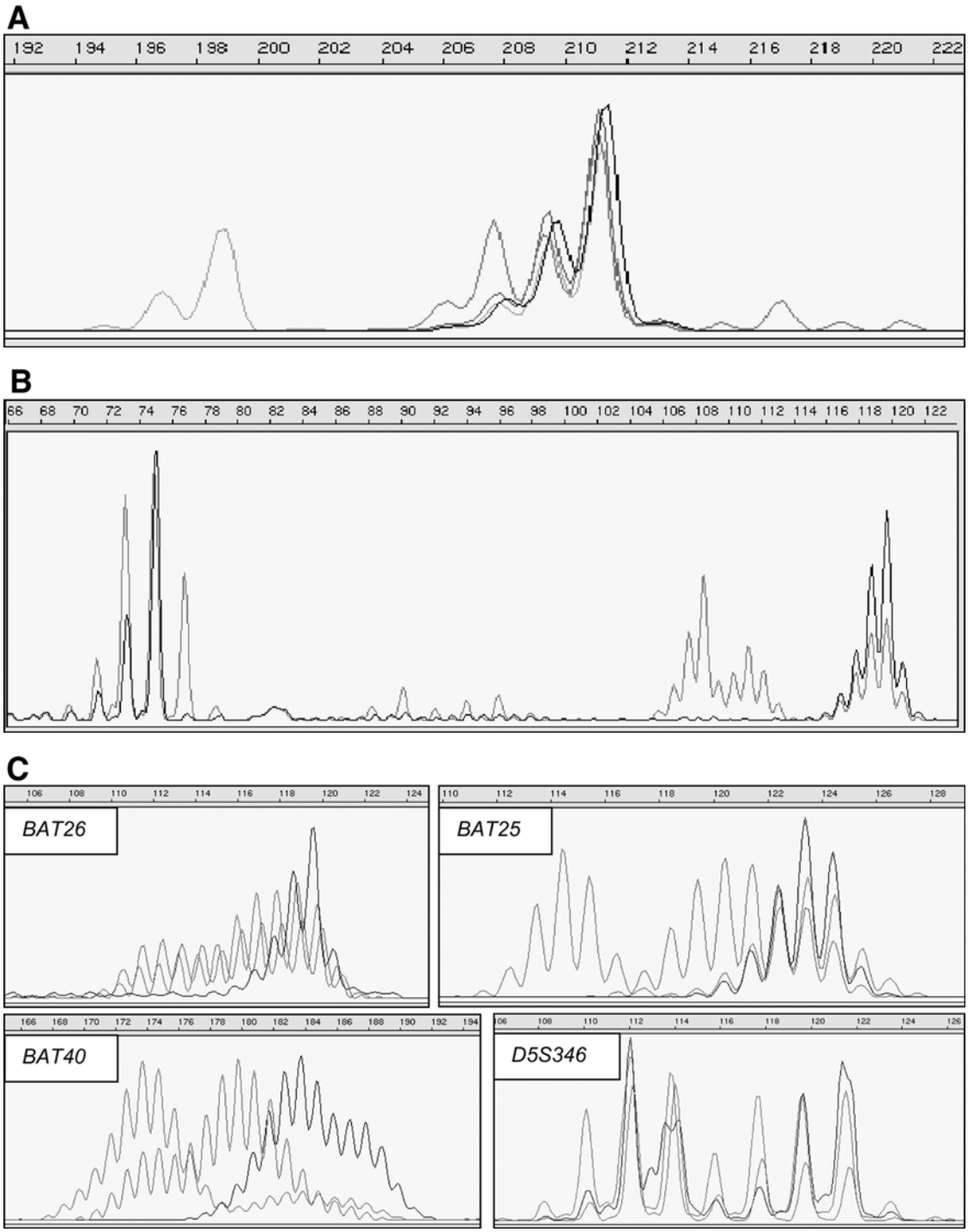


Fig. 3. Examples of microsatellite instability. **(A)** *D2S123:* Normal tissue (blood, in black, and colonic mucosa, in blue) shows a single allele at 212 bp, whereas the first tumor (red) shows instability manifest as extra alleles at 208, 218, and 222 bp, and the second (green) tumor is also unstable, with an extra allele at 200 bp. Note that if normal DNA were not available for comparison, the second (green) tumor, with two alleles, could not be scored as unstable. **(B)** Multiplex PCR: *ACTC* and *BAT26*: Normal tissue (black) from this patient shows a single allele at *ACTC* (75 bp), whereas the colon cancer (red) shows extra peaks at 77 and 88–96 bp and an increased peak height at 73 bp (which would be difficult to spot if analyzed on a conventional gel). The mononucleotide repeat (*BAT26*) shows extra alleles at 109 and 112 bp. Instability is, thus, manifest at two loci. **(C)** *BAT26, BAT25, D5S346, BAT40*: Two different regions of this tubulovillous adenoma were sampled (red and blue). Normal blood DNA is in black. Although a similar pattern of instability is observed at *BAT26* in DNA from both parts of the tumor, two distinct patterns of instability are seen at *BAT25*, *D5S346,* and *BAT40*, evidence of sub clonal evolution in the tumor. (Figure appears in color in insert following p. 172.)

As discussed in Subsection 2.2.3, the absence of MSI carries a high negative predictive value; that is, it provides good to excellent evidence that HNPCC was not the cause of the tumor tested: however, it does not necessarily exclude that individual from carrying a germline mutation in an MMR gene *(74)*. Consistent absence of MSI in tumors from the same individual does, though, substantially reduce the chances of (i.e., effectively rules out) HNPCC as a diagnosis in that individual.

Similarly, consistent absence of MSI in tumors from the same family effectively rules out HNPCC as a diagnosis in that family.

3. MMR GENE MUTATION DETECTION

As HNPCC is genetically heterogeneous, mutation detection has to encompass a number of genes. Fortunately, germline mutations in *MSH2* (2p21) and *MLH1* (3p22.3) account roughly equally for >90% HNPCC *(1,61,90–92)*. A small number of families (<5%) have mutations in *MSH6* (2p16.3; but only 300 kb from *MSH2* and *MSH6*), whereas <1% have mutations in *PMS2* (7p22.1). If *PMS1* and *MSH3* germline mutations are a cause of HNPCC, then they are very rare *(93)*. All classes of mutation are associated with HNPCC. Nonsense, frameshift, splice, and missense point mutations all occur, although an increasing number of whole exon deletions/duplications and rearrangements are being found, especially in *MSH2*, where they might account for perhaps a third or more of mutations at that locus *(94–97)*. Only one kindred has so far been described with a large chromosomal defect, a 10-Mb paracentric inversion of 2p with a breakpoint involving *MSH2 (98)*. Promoter mutations have also been described *(99)*. Thus, for complete mutation detection, both point mutations and whole-exon deletions/duplications must be covered. However, methods and techniques must take into account polymorphisms and pseudogenes *(100,101)*.

3.1. POINT MUTATION DETECTION IN HNPCC There are many different techniques that might be and have been applied to point mutation detection in HNPCC, including prescreening by a technique such as single-strand conformational polymorphism (SSCP), denaturing high-performance liquid chromatography (DHPLC) or denaturing guadrant gel electro phoresis (DGGE) prior to sequencing *(29,90–92)*. However, we use a technique that was developed in our laboratory that has considerable advantages: comparative sequence analysis (CSA) *(102,103)*. Although sequencing reactions can be analyzed conventionally, we have found that the sensitivity to detect mutations is not as high with conventional sequencing as it is with CSA, and conventional sequencing is certainly a considerably slower method of data analysis.

3.1.1. COMPARATIVE SEQUENCE ANALYSIS Comparative sequence analysis is a simple, quick, and secure method for directly comparing baselined raw data from equivalent sequencing reactions on different samples without the need for base calling. The control sequence electropherogram is split into its four component traces (one for each base), which are overlaid with equivalent traces from a second reaction. The resulting format allows simple and rapid visualization of any differences (mutations) between the two sets of traces (Fig. 4). Because the assay is based on sequencing, mutations can be characterized as well as identified in the same process. Moreover, because there has been minimal software manipulation of the data, heterozygous mutations that are difficult to pick up using standard sequencing analysis techniques have been shown to be easily detectable in a single direction using CSA. Finally, and perhaps most importantly from a diagnostic point of view, CSA is a simple and rapid technique for verifying normal sequence data with a very high degree of confidence.

To perform CSA, conventional sequencing reactions are prepared and run on an ABI fluorescent DNA analyzer (ABI 377 or 3100), but the data analysis is carried out in a different way (Fig. 4) *(102,103)*. One advantage of CSA is that PCR amplicons do not need to be kept as short as possible to maximize sensitivity in mutation detection, unlike methods requiring prescreening of an amplicon for a conformational change (e.g., SSCP). Thus, we have taken advantage of this and designed sets of primers for *MSH2* and *MLH1* that adequately cover intron–exon junctions by a wide margin and avoid known polymorphisms that might cause allele drop out. By use of generic biotinylated M13 sequencing primers, the sequencing reactions themselves are streamlined. Note that the amplicons need only be sequenced in one direction (e.g., the direction avoiding going through a repeat if that is an issue). Effectively, any one mutation is detected by at least two sequencing reactions (i.e., C and T, or G and A).

Full details of how to perform CSA are available at http://www.ngrl.co.uk/Wessex/csa.htm or from the authors (IMF). The GDB reference sequence for *MSH2* is NM_000251, and for *MLH1,* it is NM_000249 (http://www.ncbi.nlm.nih.gov./entrez/query.fcgi). To aid in mutation detection, we have found it useful to produce files showing sequences of the *MSH2* and *MLH1* PCR products and common polymorphisms by exon and translations of the cDNA sequences, also showing polymorphisms. These are also available on request from the authors (IMF).

The ICG-HNPCC database of MMR gene mutations (http://www.insight-group.org) is useful for cross-checking putative mutations. Finding a nonsense or frameshifting mutation is in itself good evidence of pathogenicity, but it is always satisfying to see this backed up by IHC data. Missense and splice-site mutations are always contentious, but if backed up by IHC data, the evidence for pathogenicity is considerably increased, especially if both major genes have been screened for mutations and the only sequence change is found in the gene predicted by IHC. It is worth checking all sequence changes, including apparently 'silent' mutations, to see if they might predict aberrant splicing. We have found GeneSplicer (http://www.tigr.org/tdb/GeneSplicer/gene_spl.html) to be particularly useful.

If laboratories wish to carry out mRNA studies on putative splice site mutations, it should be borne in mind that expression of many genes in peripheral blood leucocytes might not be representative of expression in tissues affected in that condition. In particular, mRNA splicing patterns could differ *(104,105)*. Functional studies, other than IHC, might be the only way of interpreting some putative pathogenic mutations *(40,106–108)*

3.2. WHOLE-EXON MUTATION DETECTION IN HNPCC Although CSA is excellent at finding point mutations, it cannot find larger deletions/insertions encompassing whole exons; to detect these, another strategy is required. A number of techniques exist allowing dosage of individual exons to be determined. Southern blotting has not just the potential for finding large deletions/insertions but also rearrangements. However, it is time-consuming and, hence, expensive, typically uses radionuclides, requires a large amount of DNA, and is far from being a robust technique. Fortunately, a couple of PCR-based techniques have become available that are able to address this problem.

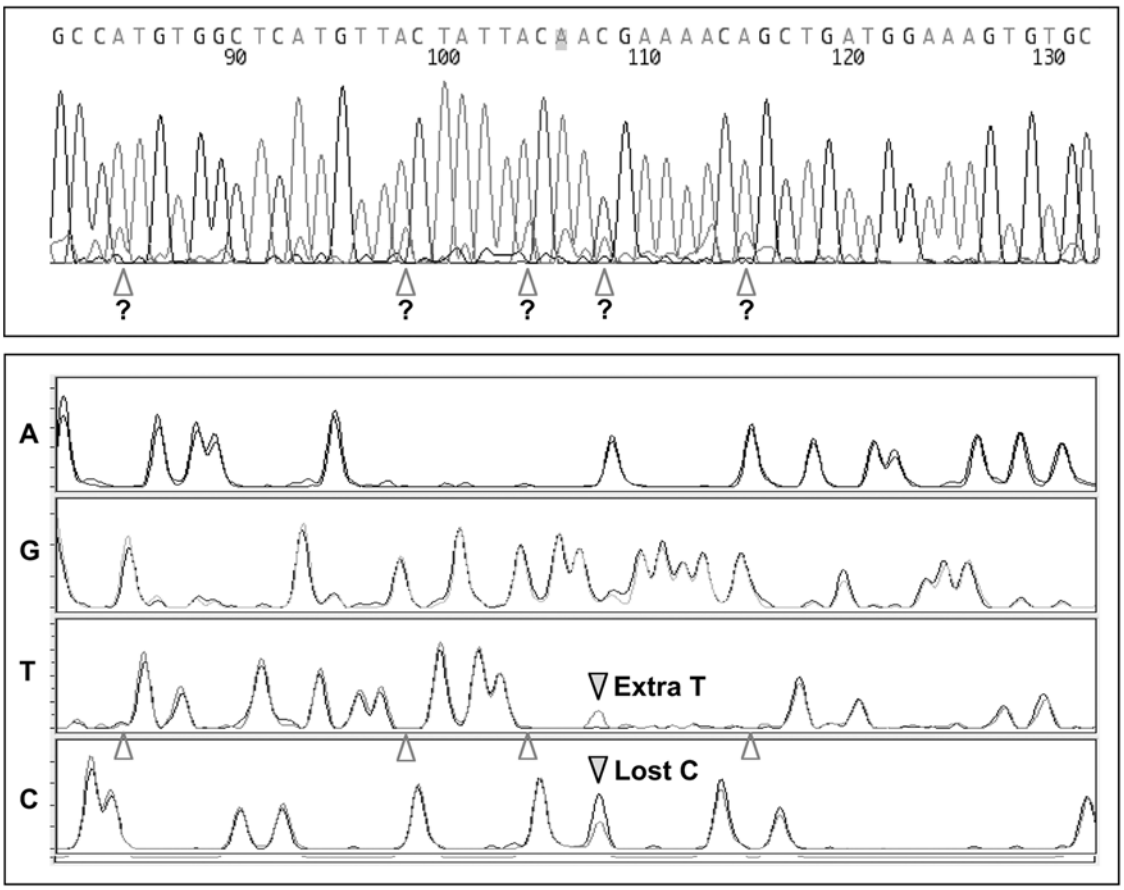


Fig. 4. Comparative sequence analysis. Conventional sequencing software output is shown in the top panel. The open red triangles indicate possible heterozygous mutations. When the file is analyzed in Genotyper, such that the four components are split (A, G, T, C), and overlaid with traces from a normal sample (black), it becomes obvious which is the true mutation (a C>T [solid green triangle]), and that the others were the result of noise. (Figure appears in color in insert following p. 172.)

Real-time QF-PCR is a possibility but requires a dedicated analyzer. Multiplex amplifiable probe hybridization (MAPH) is also a possibility *(109,110)*. However, we currently use a third technique: multiplex ligation-dependent probe amplification (MLPA). It is relatively quick, inexpensive, simple, and reproducible and uses equipment we already have for sequencing (an ABI 3100 capillary sequencer). A commercially available kit is available from MRC-Holland (http://www. mrc-holland.com/) *(111,112)*.

Multiplex ligation-dependent probe amplification is a relatively new method to detect copy number changes of genomic sequences such as single exons. Genomic DNA is denatured and a mixture of oligonucleotide probes is hybridized to the DNA. In the HNPCC MLPA kit, there are 42 probes: 7 control probes, 19 *MLH1* probes, and 16 *MSH2* probes (one for each exon). Each MLPA probe consists of two oligonucleotides, designed to hybridize to specific adjacent target sequences, which are ligated enzymatically. The lengths of the probes are all different, but the 5′ ends of all the forward primers are common, as are the 3′ ends of all the reverse primers. The ligated products are then amplified simultaneously using a pair of PCR primers complementary to the common ends of the probes, one of which is fluorescently labeled. As each probe is a different length, the amplified products are easily distinguished on a fluorescent DNA analyzer. The relative amounts of probe amplification products reflect the relative copy number of the target sequences. As a high proportion of HNPCC mutations is the result of whole-exon insertions/deletions, particularly involving *MSH2*, then MLPA is a most cost-effective contribution to mutation detection in this condition.

The exact proportion of mutations detectable by MLPA varies between laboratories, as does the relative contribution of such mutations in *MSH2* and *MLH1*. In the Netherlands, up to 50% of *MSH2* and 5% of *MLH1* mutations have been found to be the result of whole-exon insertion/deletions, whereas we have found in East Anglia that the proportions are closer to 25% for *MSH2* and 5% for *MLH1*, whereas a rather higher proportion of *MLH1* mutations are found in northern England [see 4.1.2 Mutation detection of large rearrangements, in CMGS Best Practice Guidelines: Molecular Genetics Service Testing for Hereditary Non-Polyposis Colorectal Cancer (HNPCC), at http://www.cmgs.org/BPG/Default.htm]. Undoubtedly, a combination of factors contribute to this, such as family selection and local population genetics *(94,95,97)*.

3.2.1. Multiplex Ligation-Dependent Probe Amplification: Practical Points Like any dosage technique, MLPA is particularly sensitive to DNA quality. We have observed that the occasional sample could initially give spurious results, but this can be addressed by always carrying out tests in duplicate. If a discordant result is be found the first time, repeat the analysis, perhaps including other DNA samples from the same and/or

different individuals in the same family (affected and unaffected if a binding site polymorphism is suspected) *(100)*.

Caution needs to be exercised when only a single exon appears affected, and we would not interpret such a finding as definite without corroborating evidence, such as from long-range PCR, quantitative PCR, or Southern blot. It is possible that, for example, a primer-binding site polymorphism might be responsible for the apparent 'drop out' of one exon *(100)*. This might segregate with disease in a family if it is in cis with the pathogenic mutation. This problem might also affect other PCR-based tests to diagnose copy number changes, such as reverse-transcription (RT)-PCR, although it is unlikely that both assays, MLPA and RT-PCR, would be affected. Availability of IHC data from the family is also a great help in resolving such issues, as can proceeding to full point mutation detection.

A consistent pattern on MLPA of deletion and/or insertion of more than one exon from the same gene can be interpreted as a mutation. However, if the affected exons are not contiguous, then further confirmatory tests will be required. In this respect, IHC data can be very useful.

3.3. MMR GENE MUTATION DETECTION STRATEGY

It has been mentioned earlier that only a minority of familial bowel cancer is the result of HNPCC. Moreover, although some families clearly have HNPCC, others might be indistinguishable from those with moderate risk because of other causes. It will thus be evident that although family history is a reasonable predictor of HNPCC, it has its limitations. One major problem is that people have understandably latched onto the concept that a family meeting the Amsterdam criteria equals HNPCC, and this is far from the case. It bears stating again that the Amsterdam criteria were designed to select families for research studies, not for clinical diagnosis, and were meant to be specific at the cost of sensitivity—they were also designed before the genetic basis of HNPCC was known. Wijnen and colleagues have taken a practical approach to this problem by looking for mutations in *MSH2* and *MLH1*, using an efficient technique (DGGE), in a large set of unselected families with a history of bowel cancer *(29)*. After mutation detection, they then performed multivariate analyses to determine which features of family history predicted finding a mutation, resulting in two models, P_1 and P_2, to calculate the odds of finding a mutation. Model P_1 was found to work better with medium to large families, and P_2 was found better for predicting the odds in small families (Table 9 and Fig. 5). The calculation can be facilitated by incorporation in a database or spreadsheet.

As Wijnen and colleagues point out, the utility of this approach is that families can be ranked by the odds of a mutation in order to prioritize resources for mutation detection *(29)*. Although their data did not take into account families with whole-exon deletions or insertions, undetectable by DGGE, there is no reason to believe that the phenotype of such families is different from those with point mutations. Thus, although the absolute odds of a mutation estimated in this way will be an underestimate, relative ranking of families should not be significantly affected.

Wijnen and colleagues suggested a cutoff value of 20%: above this value, families are offered mutation detection, but below it, they are not, unless there is some clinical feature not taken into account in the models that nonetheless is a good indicator of HNPCC (e.g., sebaceous skin tumors [Muir–Torre syndrome]) *(1,5,64,84,86)*. We have found that about 80% of historical referrals for HNPCC testing fall below a cutoff of 20%. Note that in cases of clinical suspicion, but odds less than 20%, it can be most useful to carry out tumor studies, especially MSI testing. The absence of MSI in a tumor carries a very high negative predictive value of HNPCC, in effect excluding it, something that no amount of negative findings on mutation detection is capable of (Table 10) *(74)*. Being able to reduce the chances of HNPCC to such low levels gives confidence to clinical decisions to treat a family as moderate risk, rather than at risk of HNPCC. Thus, prioritizing families in such a fashion does not deny them testing, but, rather, directs them to the most appropriate testing. In a situation in which a large number of families are receiving more intensive HNPCC-type colonoscopic screening programs but are only at moderate risk of bowel cancer, MSI testing can play a large part in relieving local endoscopy services. In effect, the resources saved by not carrying out a large amount of mutation detection can be directed to more appropriate tumor testing.

Our overall strategy for HNPCC testing is shown in Fig. 6. This can be altered to allow for local variation. If, for example, it is decided not to test in circumstances where the odds of a mutation in a family is less than 20%, then tumors need not be sought from such families (although we would advocate that testing in such circumstances is worthwhile if it would allow families or individuals having 2-yearly colonoscopies to be reclassified as being at moderate risk and thus only require 5-yearly colonoscopy). If it was decided not to perform MSI testing, then families with odds of >20% (or deemed to be so clinically), but with tumours with normal IHC, could proceed straight to mutation detection. If there is a backlog of families requiring germline mutation detection, then the odds of a mutation can be used to prioritize this work, unless there is some pressing clinical need in particular cases. Note that mutation detection can be limited to one gene when IHC of that protein is found to be abnormal, although it can sometimes be helpful to have tested both *MSH2* and *MLH1*.

4. MULTIDISCIPLINARY APPROACH

This approach to diagnosis necessitates the cooperation of both histopathology and molecular genetics laboratories. We would not advocate that genetics laboratories attempt IHC on their own, or indeed tumor work of any sort, without the close cooperation of a histopathologist. Similarly, although some histopathology laboratories might feel that they can attempt MSI work on their own, they would be advised to seek the close cooperation of colleagues in a diagnostic molecular genetics laboratory, which might, effectively, be mandatory under some clinico-pathological accreditation schemes. It is likely that this multidisciplinary pattern of working will become increasingly widespread as molecular pathology develops *(113,114)*.

5. IHC OR MSI?

One frequently asked question is whether MSI is better than IHC, and if a laboratory can only do one, which should it be

Table 9
Features of Family History That Are Independent Predictors of the Odds of a Mutation in *MSH2* or *MLH1*

	Variable	
Feature	*Model P_1*	*Model P_2*
Average age of onset of bowel cancer(s) in years[a]	V_1	V_1
Whether any women have developed endometrial cancer	V_2	V_3
Whether the family fulfil the modified Amsterdam criteria (as per Table 2)	V_3	
Number of cases of bowel cancer in the family		V_2

[a]Average age of onset of first bowel cancers, if individuals have developed more than one.
Probability of mutation in *MSH2* or *MLH1*:

$$p = e^L/(1 - e^L)$$

For model P_1 (larger families):

$$L = 1.4 + (-0.1)V_1 + 1.7V_2 + 2.4V_3$$

where V_1 is the mean age of colorectal cancers (yr); V_2 is the 1 if yes, else = 0; and V_3 is the 1 if AC+, else = 0.
For model P_2 (smaller families):

$$L = 1.8 + (-0.1)V_1 + 0.3V_2 + 0.8V_3$$

where V_1 is the mean age CRCa, V_2 is the number of CRCa-affected family members, and V_3 is the number of ENCa-affected family members.
Source: ref. *29.*

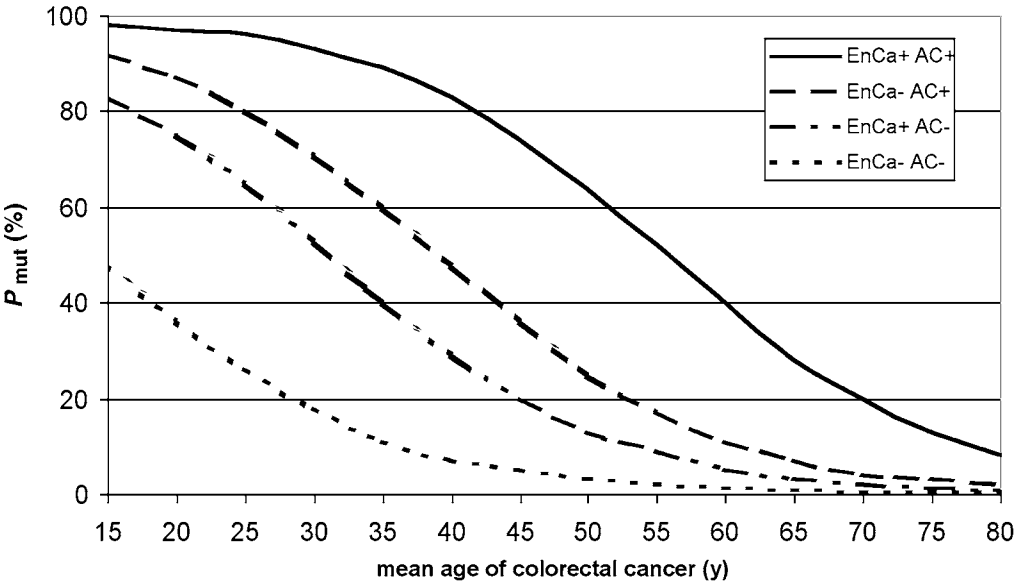


Fig. 5. Mutation probability by family type (model P_1). P_{mut} (%) is the probability of mutation in either *MSH2* or *MLH1*; AC+ is the that family fulfills the Amsterdam criteria; AC– is that the family does not fulfill the Amsterdam criteria, EnCa+ is that the family contains at least one case of endometrial cancer; EnCa– is that the family does not have a case of endometrial cancer.

(115,116)? Our answer is that IHC and MSI give complementary information, and they are not strictly comparable, so it cannot be said that one is better than the other. IHC gives more information, in that it narrows down the putative affected gene, but about one-third of tumors with MSI do not have abnormal IHC, so not finding abnormal IHC does not add any information for or against HNPCC. However, finding abnormal IHC does give considerable information. Given the relatively high rate of sporadic MLH1 loss in colon cancers, finding an *MSH2* abnormality carries a higher predictive value than an *MLH1* abnormality. Finding any IHC abnormality in a rectal cancer, or colorectal adenoma, is significant. In contrast, MSI indicates an MMR abnormality, but not, of course, which gene. It takes somewhat more work to perform compared with IHC, but the negative predictive value of not finding MSI is substantial *(74)*. Thus, in terms of one being better than the other; IHC is better if one wishes to include HNPCC and it is likely to be found (i.e., in cases where the odds of a mutation are >20%), whereas MSI is better if one wishes to exclude HNPCC and it is unlikely to be found (i.e., when the odds of a mutation are <20%). As we have yet to find a tumor with abnormal IHC that does not have MSI, then we would not test for MSI on tumors from families with high odds of a mutation, but we would perform IHC first and then MSI if the IHC turns out to be normal (Figs. 3 and 6).

6. COSTS

A number of studies have addressed the cost-effectiveness of tumor testing in HNPCC *(116–119)*. [Of the value of

Table 10
Positive and Negative Predictive Values of MSI in Colon tumors

	MSI detected in			
	One Cancer	*One Adenoma*	*Two Cancers*	*Two Adenomas*
Sensitivity	95%	95%	90%	81%
Specificity	86%	98%	98%	>99%
Positive predictive value	6%	32%	29%	95%
Negative predictive value	>99%	>99%	98%	>99%

[a]Note that these figures relate to colon, tumors, not rectal tumors. Rectal tumors with MSI carry a high positive predictive value of HNPCC *(83)*.
Source: ref. *73*.

colonoscopic surveillance in those at-risk of HNPCC, there is now little doubt *(120)*.] Based on our experience, modelling (Table 11) shows that because MLPA is substantially less expensive to perform than CSA, financial savings can be made by carrying out MLPA first, in spite of the fact that it tests both *MSH2* and *MLH1* and thus information from IHC cannot be used to limit it to one gene. When IHC is carried out, CSA can be limited to one gene in approximately two-thirds of cases, enabling even greater savings. Although it might appear at first sight that the extra time and effort involved in carrying out IHC is not worth the financial gain, account must be taken of the fact that IHC data enables a higher proportion of mutations to be interpreted as pathogenic. When this is taken into account, the costs per interpretable mutation are substantially less when IHC is performed (Table 12). We have found that approx 60% of HNPCC-associated mutations predict either missense amino acid or splice

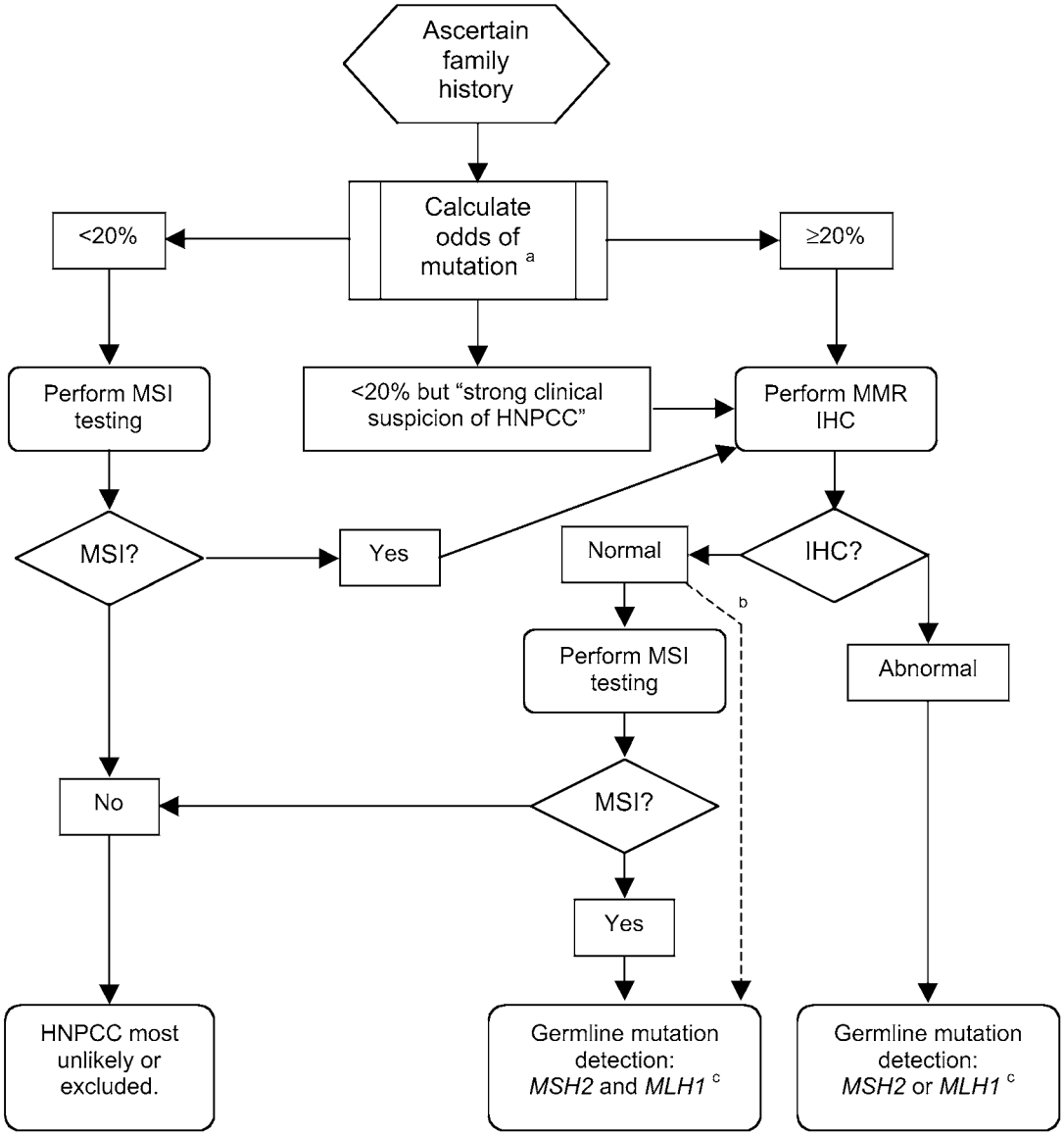


Fig. 6. Overall HNPCC testing strategy. [a]Calculated according to Wijnen et al. *(29)*. [b]Laboratories not wishing to perform MSI testing can follow the dashed arrow on finding normal MMR expression on IHC. [c]In rare cases, mutation detection of *MSH6* or *PMS2* might be appropriate.

Table 11
Cost Modeling of HNPCC Mutation Detection Strategies: MLPA Before CSA and Vice Versa, With or Without Taking into Account IHC Data

	MLPA first / CSA second				*CSA first / MLPA second*			
	Test	*Cases*	*Cost per test*	*Total*	*Test*	*Cases*	*Cost per test*	*Total*
Without IHC data	MLPA	100	£10	£1,000	CSA	100	£400	£40,000
	CSA	85[a]	£400	£34,000	MLPA	15	£10	£150
				£35,000				£40,150
With IHC data	IHC	100	£75	£7,500	IHC	100	£75	£7,500
	MLPA	100	£10	£1,000	CSA	67	£200	£13,400
	CSA	28[b]	£400	£11,200	CSA	33	£400	£13,200
	CSA	57	£200	£11,400	MLPA	15	£10	£150
				£31,100				£34,250

[a]Approximately 15% of mutations are detectable by MLPA
[b]Approximately 33% of mutations are not associated with abnormal IHC; CSA of one gene costs £200, and two genes cost £400.

Table 12
Cost Modeling of HNPCC Mutation Detection Strategies: The Cost of Finding Mutations with Interpretability Factored In

	Mutations	*Total cost*[a]	*Cost/ mutation*	*Interpretable mutations*	*Cost/ mutation*
No IHC	58[b]	£35,000	£603	23	£1,522
IHC	58	£31,100	£536	49	£635

[a]From Table 11.
[b]Number of mutations expected after testing 100 families with odds of mutation between 20% and 100%.

site changes; that is, they would be difficult or impossible to interpret without some other independent evidence such as from IHC. Thus, by performing IHC a service genetics laboratory can minimize the number of inconclusive mutation detection reports it must issue with the caveat that this mutation is of uncertain significance.

7. OVERALL INTERPRETATION OF TEST RESULTS

Individual laboratories will want to use their own reporting style and standards and determine what local strategy for HNPCC testing they wish to carry out (e.g., whether to carry out just IHC and mutation detection, or include MSI testing). HNPCC best practice guidelines for UK clinical molecular genetic laboratories can be found at http://www.cmgs.org/BPG/Default.htm, where guidelines for a range of other techniques and processes can also be found.

Rather than exclusively comment on the risk of cancer in a report, it might be better to discuss an individual's risk of an "HNPCC-associated tumor," which will cover, for example, colorectal adenomas and skin lesions. Because knowledge of HNPCC is incomplete, it is better to keep an open mind *(121).* Although an individual's risk of HNPCC might be reduced by genetic testing, his risk of cancer might not necessarily be reduced, given that a family history of cancer carries with it an empiric raised risk, only part of which is the result of HNPCC *(122).*

ACKNOWLEDGMENTS

We would like to acknowledge the counsel and assistance of Rachel Butler, Ruth Charlton, Jeremy Jass, Peter Laird, Gabriela Möslein, Sheila Palmer-Smith, George Poulogiannis, Melissa Southey, Patrick Tarpey, Ian Tomlinson, Juul Wijnen, and Nikki Wood.

REFERENCES

1. Lynch, H. T., and de la Chapelle, A. Hereditary colorectal cancer. *N. Engl. J. Med.* **348:**919–932, 2003.
2. Warthin, A. S. Heredity with reference to carcinoma, as shown by the study of the cases examined in the pathological laboratory of the University of Michigan, 1895–1913. *Arch. Intern. Med.* **12:** 546–555, 1913.
3. Lynch, H. T., Shaw, M. W., Magnuson, C. W., Larsen, A. L., and Krush, A. J. Hereditary factors in cancer: study of two large midwestern kindreds. *Arch. Intern. Med.* **177:**206–212, 1966.
4. Lynch, H. T., Kimberling, W., Albano, W. A., et al. Hereditary nonpolyposis colorectal cancer (Lynch syndromes I and II). I. Clinical description of resource. *Cancer* **56:**934–938, 1985.
5. Frayling, I. M. Hereditary non-polyposis colorectal cancer (HNPCC), in *Oxford Desk Reference—Clinical Genetics,* Firth, H. V., and Hirst, J. A., eds., Hall, J. G., consulting ed., Oxford University Press, Oxford, 2005.
6. Jass, J. R. Familial colorectal cancer: pathology and molecular characteristics. *Lancet Oncol.* **1:**220–226, 2000.
7. Johns, L. E., and Houlston, R.S. A systematic review and meta-analysis of familial colorectal cancer risk. *Am. J. Gastroenterol.* **96:**2992–3003, 2001.
8. Evans, D. G., Walsh, S., Jeacock, J., et al. Incidence of hereditary non-polyposis colorectal cancer in a population-based study of 1137 consecutive cases of colorectal cancer. *Br. J. Surg.* **84:**1281–1285, 1997.
9. Katballe, N., Christensen, M., Wikman, F. P., Orntoft, T. F., and Laurberg, S. Frequency of hereditary non-polyposis colorectal cancer in Danish colorectal cancer patients. *Gut* **50:**43–51, 2002.
10. Samowitz, W. S., Curtin, K., Lin, H. H., et al. The colon cancer burden of genetically defined hereditary nonpolyposis colon cancer. *Gastroenterology* **121:**830–838, 2001.
11. Salovaara, R., Loukola, A., Kristo, P., et al. Population-based molecular detection of hereditary nonpolyposis colorectal cancer. *J. Clin. Oncol.* **18:**2193–2200, 2000.
12. Peel, D. J., Ziogas, A., Fox, E. A., et al. Characterization of hereditary nonpolyposis colorectal cancer families from a population-based series of cases. *J. Natl. Cancer Inst.* **92:**1517–1522, 2000.
13. Foulkes, W. D., Thiffault, I., Gruber, S. B., et al. The founder mutation MSH2*1906G → C is an important cause of hereditary nonpolyposis colorectal cancer in the Ashkenazi Jewish population. *Am. J. Hum. Genet.* **71:**1395–1412, 2002.
14. Nystrom-Lahti, M., Kristo, P., Nicolaides, N. C., et al. Founding mutations and Alu-mediated recombination in hereditary colon cancer. *Nat. Med.* **1:**1203–1206, 1995.
15. Vasen, H. F., Watson, P., Mecklin, J. P., and Lynch, H.T. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. *Gastroenterology* **116:**1453–1456, 1999.

16. Houlston, R. S., and Tomlinson, I.P. Polymorphisms and colorectal tumor risk. *Gastroenterology* **121:**282–301, 2001.
17. Lynch, H. T., Smyrk, T., and Lynch, J. F. Molecular genetics and clinical-pathology features of hereditary nonpolyposis colorectal carcinoma (Lynch syndrome): historical journey from pedigree anecdote to molecular genetic confirmation. *Oncology* **55:**103–108, 1998.
18. Fishel, R., Lescoe, M. K., Rao, M. R., et al. The human mutator gene homolog, MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell* **75:**1027–1038, 1993; erratum: *Cell* **77:**167, 1994.
19. Hall, N. R., Taylor, G., Finan, P., et al. Intron splice acceptor site sequence variation in the hereditary non-polyposis colorectal cancer gene hMSH2. *Eur. J. Cancer* **30A:**1550–1552, 1994.
20. Bronner, C. E., Baker, S. M., Morrison, P. T., et al. Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature* **368:**258–261, 1994.
21. Lothe, R. A., Peltomaki, P., Meling, G. I., et al. Genomic instability in colorectal cancer: relationship to clinicopathological variables and family history. *Cancer Res.* **53:**5849–5852, 1993.
22. Parsons, R., Li, G. M., Longley, M. J., et al. Hypermutability and mismatch repair deficiency in RER+ tumor cells. *Cell* **75:**1227–1236, 1993.
23. Peltomaki, P., Lothe, R. A., Aaltonen, L. A., et al. Microsatellite instability is associated with tumors that characterize the hereditary non-polyposis colorectal carcinoma syndrome. *Cancer Res*. **53:**5853–5855, 1993.
24. Thibodeau, S. N., French, A. J., Roche, P. C., et al. Altered expression of hMSH2 and hMLH1 in tumors with microsatellite instability and genetic alterations in mismatch repair genes. *Cancer Res.* **56:**4836–4840, 1996.
25. Kolodner, R. D., Hall, N. R., Lipford, J., P et al. Structure of the human MSH2 locus and analysis of two Muir-Torre kindreds for msh2 mutations. *Genomics* **24:**516–526, 1994.
26. Muller, A., and Fishel, R. Mismatch repair and the hereditary nonpolyposis colorectal cancer syndrome (HNPCC). *Cancer Invest*. **20:**102–109, 2002.
27. Peltomaki, P. Role of DNA mismatch repair defects in the pathogenesis of human cancer. *J. Clin. Oncol*. **21:**1174–1179, 2003.
28. Aaltonen, L. A., Peltomaki, P., Mecklin, J. P., et al. Replication errors in benign and malignant tumors from hereditary nonpolyposis colorectal cancer patients. *Cancer Res.* **54:**1645–1648, 1994.
29. Wijnen, J. T., Vasen, H. F., Khan, P. M., et al. Clinical findings with implications for genetic testing in families with clustering of colorectal cancer. *N. Engl. J. Med.* **339:**511–518, 1998.
30. de Leeuw, W. J., Dierssen, J., Vasen, H. F., et al. Prediction of a mismatch repair gene defect by microsatellite instability and immunohistochemical analysis in endometrial tumours from HNPCC patients. *J. Pathol.* **192:**328–335, 2000.
31. Debniak, T., Kurzawski, G., Gorski, B., Kladny, J., Domagala, W., and Lubinski, J. Value of pedigree/clinical data, immunohistochemistry and microsatellite instability analyses in reducing the cost of determining hMLH1 and hMSH2 gene mutations in patients with colorectal cancer. *Eur. J. Cancer.* **36:**49–54, 2000.
32. Ruszkiewicz, A., Bennett, G., Moore, J., et al. Correlation of mismatch repair genes immunohistochemistry and microsatellite instability status in HNPCC-associated tumours. *Pathology* **34:**541–547, 2002.
33. Salahshor, S., Koelble, K., Rubio, C., and Lindblom, A. Microsatellite Instability and hMLH1 and hMSH2 expression analysis in familial and sporadic colorectal cancer. *Lab. Invest.* **81:** 535–541, 2001.
34. Wahlberg, S. S., Schmeits, J., Thomas, G., et al. Evaluation of microsatellite instability and immunohistochemistry for the prediction of germ-line MSH2 and MLH1 mutations in hereditary nonpolyposis colon cancer families. *Cancer Res*. **62:**3485–3492, 2002.
35. Ward, R., Meldrum, C., Williams, R., et al. Impact of microsatellite testing and mismatch repair protein expression on the clinical interpretation of genetic testing in hereditary non-polyposis colorectal cancer. *J. Cancer Res. Clin. Oncol.* **128:**403–411, 2002.
36. Christensen, M., Katballe, N., Wikman, F., et al. Antibody-based screening for hereditary nonpolyposis colorectal carcinoma compared with microsatellite analysis and sequencing. *Cancer* **95:** 2422–2430, 2002.
37. Genschel, J., Littman, S. J., Drummond, J. T., and Modrich, P. Isolation of MutSbeta from human cells and comparison of the mismatch repair specificities of MutSbeta and MutSalpha. *J. Biol. Chem*. **273:**19,895–19,901, 1998.
38. Palombo, F., Iaccarino, I., Nakajima, E., Ikejima, M., Shimada, T., and Jiricny, J. hMutSbeta, a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA. *Curr. Biol.* **6:**1181–1184, 1996.
39. Raschle, M., Marra, G., Nystrom-Lahti, M., Schar, P., and Jiricny, J. Identification of hMutLbeta, a heterodimer of hMLH1 and hPMS1. *J. Biol. Chem.* **274:**32,368–32,375, 1999.
40. Guerrette, S., Wilson, T., Gradia, S., and Fishel, R. Interactions of human hMSH2 with hMSH3 and hMSH2 with hMSH6: examination of mutations found in hereditary nonpolyposis colorectal cancer. *Mol. Cell. Biol*. **18:**6616–6623, 1998.
41. Guerrette, S., Acharya, S., and Fishel, R. The interaction of the human MutL homologues in hereditary nonpolyposis colon cancer. *J. Biol. Chem*. **274:**6336–6341, 1999.
42. Heinen, C. D., Wilson, T., Mazurek, A., Berardini, M., Butz, C., and Fishel, R. HNPCC mutations in hMSH2 result in reduced hMSH2-hMSH6 molecular switch functions. *Cancer Cell* **1:**469–478, 2002.
43. Kariola, R., Raevaara, T. E., Lonnqvist, K. E., and Nystrom-Lahti, M. Functional analysis of MSH6 mutations linked to kindreds with putative hereditary non-polyposis colorectal cancer syndrome. *Hum. Mol. Genet.* **11:**1303–1310, 2002.
44. Plaschke, J., Kruger, S., Pistorius, S., Theissig, F., Saeger, H. D., and Schackert, H. K. Involvement of hMSH6 in the development of hereditary and sporadic colorectal cancer revealed by immunostaining is based on germline mutations, but rarely on somatic inactivation. *Int. J. Cancer* **97:**643–648.
45. Wagner, A., Hendriks, Y., Meijers-Heijboer, E. J., et al. Atypical HNPCC owing to MSH6 germline mutations: analysis of a large Dutch pedigree. *J. Med. Genet.* **38:**318–322, 2001.
46. Wijnen, J., de Leeuw, W., Vasen, H., et al. Familial endometrial cancer in female carriers of MSH6 germline mutations. *Nat. Genet.* **23:**142–144, 1999.
47. Plotz, G., Raedle, J., Brieger, A., Trojan, J., and Zeuzem, S. N-terminus of hMLH1 confers interaction of hMutLalpha and hMutLbeta with hMutSalpha. *Nucleic Acids Res.* **31:**3217–3226, 2003.
48. Jass, J. R. hMLH1 and hMSH2 immunostaining in colorectal cancer. *Gut* **47:**315–316, 2000.
49. Jiménez-Liñán, M., Frayling, I. M., Happerfield, L., Wood, N., and Arends, M. J. Immunohistochemical expression patterns of MSH2, MLH1, MSH6 and PMS2 in normal tissues and tumours from individuals with known or suspected hereditary non-polyposis colorectal cancer. *J. Pathol.* submitted.
50. Manavis, J., Gilham, P., Davies, R., and Ruszkiewicz, A. The immunohistochemical detection of mismatch repair gene proteins (MLH1, MSH2, MSH6, and PMS2): practical aspects in antigen retrieval and biotin blocking protocols. *Appl. Immunohistochem. Mol. Morphol.* **11:**73–77, 2003.
51. Rigau, V., Sebbagh, N., Olschwang, S., et al. Microsatellite instability in colorectal carcinoma. The comparison of immunohistochemistry and molecular biology suggests a role for hMSH6 immunostaining. *Arch. Pathol. Lab. Med.* **127:**694–700, 2003.
52. Bocker, T., Diermann, J., Friedl, W., et al. Microsatellite instability analysis: a multicenter study for reliability and quality control. *Cancer Res.* **57:**4739–4743, 1997.
53. Müller, W., Burgart, L. J., Krause-Paulus, R., et al. The reliability of immunohistochemistry as a prescreening method for the diagnosis of hereditary nonpolyposis colorectal cancer (HNPCC)—Results of an international collaborative study. *Famili Cancer* **1:** 87–93, 2001.
54. Andrew, S. E., Reitmair, A. H., Fox, J., et al. Base transitions dominate the mutational spectrum of a transgenic reporter gene in MSH2 deficient mice. *Oncogene* **15:**123–129, 1997.

55. Brueckl, W. M., Jung, A., Wein, A., et al. Microsatellite instability in colorectal adenomas: relevance and clinical importance. *Int. J. Colorectal Dis*. **15:**189–196, 2000.
56. Frazier, M. L., Sinicrope, F. A., Amos, C. I., et al. Loci for efficient detection of microsatellite instability in hereditary non-polyposis colorectal cancer. *Oncol. Rep*. **6:**497–505, 1999.
57. Honchel, R., Halling, K. C., Schaid, D. J., Pittelkow, M., and Thibodeau, S. N. Microsatellite instability in Muir–Torre syndrome. *Cancer Res*. **54:**1159–1163, 1994.
58. Jass, J. R., Cottier, D. S., Jeevaratnam, P., et al. Diagnostic use of microsatellite instability in hereditary non-polyposis colorectal cancer. *Lancet* **346:**1200–1201, 1996.
59. Kruse, R., Rutten, A., Schweiger, N., et al. Frequency of microsatellite instability in unselected sebaceous gland neoplasias and hyperplasias. *J. Invest. Dermatol.* **120:**858–864, 2003.
60. Kuismanen, S. A., Moisio, A. L., Schweizer, P., et al. Endometrial and colorectal tumors from patients with hereditary nonpolyposis colon cancer display different patterns of microsatellite instability. *Am. J. Pathol*. **160:**1953–1958, 2002.
61. Liu, B., Farrington, S. M., Petersen, G. M., et al. Genetic instability occurs in the majority of young patients with colorectal cancer. *Nat. Med*. **1:**348–352, 1995.
62. Miturski, R., Bogusiewicz, M., Tarkowski, R., et al. BAT-26 microsatellite instability does not correlate with the loss of hMLH1 and hMSH2 protein expression in sporadic endometrial cancers. *Oncol. Rep*. **10:**1039–1043, 2003.
63. Stone, J. G., Tomlinson, I. P., and Houlston, R. S. Optimising methods for determining RER status in colorectal cancers. *Cancer Lett.* **149:**15–20, 2000.
64. Swale, V. J., Quinn, A. G., Wheeler, J. M., et al. Microsatellite instability in benign skin lesions in hereditary non-polyposis colorectal cancer syndrome. *J. Invest. Dermatol.* **113:**901–905, 1999.
65. Boland, C. R., Thibodeau, S. N., Hamilton, S. R., et al. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.* **58:**5248–5257, 1998.
66. Halford, S., Sasieni, P., Rowan, A., et al. Low-level microsatellite instability occurs in most colorectal cancers and is a nonrandomly distributed quantitative trait. *Cancer Res*. **62:**53–57, 2002.
67. Laiho, P., Launonen, V., Lahermo, P., et al. Low-level microsatellite instability in most colorectal carcinomas. *Cancer Res.* **62:** 1166–1170, 2002.
68. Dietmaier, W., Wallinger, S., Bocker, T., Kullmann, F., Fishel, R., and Ruschoff, J. Diagnostic microsatellite instability: definition and correlation with mismatch repair protein expression. *Cancer Res.* **57:**4749–4756, 1997.
69. Loukola, A., Eklin, K., Laiho, P., et al. Microsatellite marker analysis in screening for hereditary nonpolyposis colorectal cancer (HNPCC). *Cancer Res*. **61:**4545–4549, 2001.
70. Parsons, R., Myeroff, L. L., Liu, B., et al. Microsatellite instability and mutations of the transforming growth factor beta type II receptor gene in colorectal cancer. *Cancer Res.* **55:**5548–5550, 1995.
71. Sutter, C., Gebert, J., Bischoff, P., Herfarth, C., and von Knebel Doeberitz, M. Molecular screening of potential HNPCC patients using a multiplex microsatellite PCR system. *Mol. Cell Probes* **13:**157–165, 1999.
72. Spirio, L., Joslyn, G., Nelson, L., Leppert, M., and White, R. A CA repeat 30–70 KB downstream from the adenomatous polyposis coli (APC) gene. *Nucleic Acids Res.* **19:**6348, 1991; erratum *Nucleic Acids Res.* **20:**642, 1992.
73. Weber, J. L., Kwitek, A. E., May, P. E., et al. Dinucleotide repeat polymorphisms at the D17S250 and D17S261 loci. *Nucleic Acids Res.* **18:**4640, 1990.
74. Frayling, I. M. Microsatellite Instability. *Gut* **45:**1–4, 1999.
75. Cunningham, J. M., Christensen, E. R., Tester, D. J., et al. Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability. *Cancer Res.* **58:**3455–3460, 1998.
76. Esteller, M., Levine, R., Baylin, S. B., Ellenson, L. H., and Herman, J. G. MLH1 promoter hypermethylation is associated with the microsatellite instability phenotype in sporadic endometrial carcinomas. *Oncogene* **17:**2413–2417, 1998.
77. Yuen, S. T., Chan, T. L., Ho, J. W., et al. Germline, somatic and epigenetic events underlying mismatch repair deficiency in colorectal and HNPCC-related cancers. *Oncogene* **21:**7585–7592, 2002.
78. Tomlinson, I. P., Ilyas, M., and Bodmer, W. F. Allele loss occurs frequently at hMLH1, but rarely at hMSH2, in sporadic colorectal cancers with microsatellite instability. *Br. J. Cancer* **74:**1514–1517, 1998.
79. Esteller, M., Fraga, M. F., Guo, M., et al. DNA methylation patterns in hereditary human cancers mimic sporadic tumorigenesis. *Hum. Mol. Genet*. **10:**3001–3007, 2001.
80. Wheeler, J. M., Loukola, A., Aaltonen, L. A., Mortensen, N. J., and Bodmer, W. F. The role of hypermethylation of the hMLH1 promoter region in HNPCC versus MSI+ sporadic colorectal cancers. *J. Med. Genet.* **37:**588–592, 2000.
81. Deng, G., Chen, A., Hong, J., Chae, H. S., and Kim, Y. S. Methylation of CpG in a small region of the hMLH1 promoter invariably correlates with the absence of gene expression. *Cancer Res.* **59:**2029–2033, 1999.
82. Menigatti, M., Di Gregorio, C., Borghi, F., et al. Methylation pattern of different regions of the MLH1 promoter and silencing of gene expression in hereditary and sporadic colorectal cancer. *Genes Chromosomes Cancer* **31:**357–361, 2001.
83. Xiong, Z., Wu, A. H., Bender, C. M., et al. Mismatch repair deficiency and CpG island hypermethylation in sporadic colon adenocarcinoma *Cancer Epidemiol. Biomarkers Prev.* **10:**799–803, 2001.
84. Nilbert, M., Planck, M., Fernebro, E., Borg, A., and Johnson, A. MSI is rare in rectal carcinomas and signifies hereditary cancer. *Eur. J. Cancer* **35:**942–945, 1999.
85. Iino, H., Simms, L., Young, J., et al. DNA microsatellite instability and mismatch repair protein loss in adenomas presenting in hereditary non-polyposis colorectal cancer. *Gut* **47:**37–42, 2000.
86. Loukola, A., Salovaara, R., Kristo, P., et al. Microsatellite instability in adenomas as a marker for hereditary nonpolyposis colorectal cancer. *Am. J. Pathol*. **155:**1849–1853, 1999.
87. Pyatt, R., Chadwick, R. B., Johnson, C. K., et al. Polymorphic variation at the BAT-25 and BAT-26 loci in individuals of African origin. Implications for microsatellite instability testing. *Am. J. Pathol*. **155:**349–353, 1999.
88. Ichikawa, A., Sugano, K., and Fujita, S. DNA variants of BAT-25 in Japanese, a locus frequently used for analysis of microsatellite instability. *Jpn. J. Clin. Oncol.* **31:**346–348, 2001.
89. Samowitz, W. S., Slattery, M. L., Potter, J. D., and Leppert, M. F. BAT-26 and BAT-40 instability in colorectal adenomas and carcinomas and germline polymorphisms. *Am. J. Pathol.* **154:**1637–1641, 1999.
90. Liu, B., Parsons, R., Papadopoulos, N., et al. Analysis of mismatch repair genes in hereditary non-polyposis colorectal cancer patients. *Nat. Med*. **2:**169–174, 1996.
91. Liu, B., Parsons, R. E., Hamilton, S. R., et al. hMSH2 mutations in hereditary nonpolyposis colorectal cancer kindreds. *Cancer Res.* **54:**4590–4594, 1994.
92. Peltomaki, P., and Vasen, H. F. Mutations predisposing to hereditary nonpolyposis colorectal cancer: database and results of a collaborative study. The International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer. *Gastroenterology* **113:** 1146–1158, 1997.
93. Hienonen, T., Laiho, P., Salovaara, R., et al. Little evidence for involvement of MLH3 in colorectal cancer predisposition. *Int. J. Cancer* **106:**292–296, 2003.
94. Wijnen, J., van der Klift, H., Vasen, H., et al. MSH2 genomic deletions are a frequent cause of HNPCC. *Nat. Genet.* **20:**326–328, 1998.
95. Wang, Y., Friedl, W., Lamberti, C., et al. Hereditary nonpolyposis colorectal cancer: frequent occurrence of large genomic deletions in MSH2 and MLH1 genes. *Int. J. Cancer* **103:**636–641, 2003.
96. Plaschke, J., Ruschoff, J., and Schackert, H. K. Genomic rearrangements of hMSH6 contribute to the genetic predisposition in suspected hereditary non-polyposis colorectal cancer syndrome. *J. Med. Genet.* **40:**597–600, 2003.

97. Wagner, A., Barrows, A., Wijnen, J. T., et al. Molecular analysis of hereditary nonpolyposis colorectal cancer in the United States: high mutation detection rate among clinically selected families and characterization of an American founder genomic deletion of the MSH2 gene. *Am. J. Hum. Genet.* **72:**1088–1100, 2003.
98. Wagner, A., van der Klift, H., Franken, P., et al. A 10-Mb paracentric inversion of chromosome arm 2p inactivates MSH2 and is responsible for hereditary nonpolyposis colorectal cancer in a North-American kindred. *Genes Chromosomes Cancer* **35:**49–57, 2002.
99. Shin, K. H., Shin, J. H., Kim, J. H., and Park, J. G. Mutational analysis of promoters of mismatch repair genes hMSH2 and hMLH1 in hereditary nonpolyposis colorectal cancer and early onset colorectal cancer patients: identification of three novel germline mutations in promoter of the hMSH2 gene. *Cancer Res*. **62:**38–42, 2002. erratum *Cancer Res.* **62:**2445, 2002.
100. Andrew, S. E., Whiteside, D., Buzin, C., Greenberg, C., and Spriggs, E. An intronic polymorphism of the hMLH1 gene contributes toward incomplete genetic testing for HNPCC. *Genet. Test.* **6:**319–322, 2002.
101. Chadwick, R. B., Meek, J. E., Prior, T. W., Peltomaki, P., and de La Chapelle, A. Polymorphisms in a pseudogene highly homologous to PMS2. *Hum. Mutat.* **16:**530, 2000.
102. Mattocks, C., Tarpey, P., Bobrow, M., and Whittaker, J. Comparative sequence analysis (CSA): a new sequence-based method for the identification and characterization of mutations in DNA. *Hum. Mutat.* **16:**437–443, 2000.
103. Mattocks, C. J., Tarpey, P., and Whittaker, J. L. Comparative sequence analysis, in *Molecular Diagnosis of Genetic Diseases,* 2nd edi, Elles, R. and Mountford, R., eds., Humana, Totowa, NJ, 2003.
104. Nystrom-Lahti, M., Holmberg, M., Fidalgo, P., et al. Missense and nonsense mutations in codon 659 of MLH1 cause aberrant splicing of messenger RNA in HNPCC kindreds. *Genes Chromosomes Cancer* **26:**372–375, 1999.
105. Palmirotta, R., Veri, M. C., Curia, M. C., et al. Transcripts with splicings of exons 15 and 16 of the hMLH1 gene in normal lymphocytes: implications in RNA-based mutation screening of hereditary non-polyposis colorectal cancer. *Eur. J. Cancer* **34:**927–930, 1998.
106. Nystrom-Lahti, M., Perrera, C., Raschle, M., et al. Functional analysis of MLH1 mutations linked to hereditary nonpolyposis colon cancer. *Genes Chromosomes Cancer* **33:**160–167, 2002.
107. Raevaara, T. E., Vaccaro, C., Abdel-Rahman, W. M., et al. Pathogenicity of the hereditary colorectal cancer mutation hMLH1 del616 linked to shortage of the functional protein. *Gastroenterology* **125:**501–509, 2003.
108. Papadopoulos, N., Leach, F. S., Kinzler, K. W., and Vogelstein, B. Monoallelic mutation analysis (MAMA) for identifying germline mutations. *Nat. Genet.* **11:**99–102, 1995.
109. Hollox, E. J., Akrami, S. M., and Armour, J. A. DNA copy number analysis by MAPH: molecular diagnostic applications. *Expert Rev. Mol. Diagn.* **2:**370–378, 2002.
110. Akrami, S. M., Dunlop, M. G., Farrington, S. M., et al. Screening for exonic copy number mutations at MSH2 and MLH1 by MAPH. Familial Cancer, in press, 2005.
111. Gille, J. J., Hogervorst, F. B., Pals, G., et al. Genomic deletions of MSH2 and MLH1 in colorectal cancer families detected by a novel mutation detection approach. *Br. J. Cancer* **87:**892–897, 2002.
112. Schouten, J. P., McElgunn, C. J., et al. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* **30:**e57, 2002.
113. Tomlinson, I. P., and Ilyas, M. Molecular pathology of solid tumours: some practical suggestions for translating research into clinical practice. *Mol. Pathol.* **54:**203–205, 2001.
114. Frayling, I. M. Methods of molecular analysis: mutation detection in solid tumours. *Mol. Pathol.* **55:**73–79, 2002.
115. Lindor, N. M., Burgart, L. J., Leontovich, O., et al. Immunohistochemistry versus microsatellite instability testing in phenotyping colorectal tumors. *J. Clin. Oncol*. **20:**1043–1048, 2002.
116. Cawkwell, L., Gray, S., Murgatroyd, H., et al. Choice of management strategy for colorectal cancer based on a diagnostic immunohistochemical test for defective mismatch repair. *Gut* **45:**409–415, 1999.
117. Ramsey, S. D., Clarke, L., Etzioni, R., Higashi, M., Berry, K., and Urban, N. Cost-effectiveness of microsatellite instability screening as a method for detecting hereditary nonpolyposis colorectal cancer. *Ann. Intern. Med*. **135:**577–588, 2001.
118. Reyes, C. M., Allen, B. A., Terdiman, J. P., and Wilson, L. S. Comparison of selection strategies for genetic testing of patients with hereditary nonpolyposis colorectal carcinoma: effectiveness and cost-effectiveness. *Cancer* **95:**1848–1856, 2002.
119. Terdiman, J. P., Gum, J. R., Jr. Conrad, P. G., et al. Efficient detection of hereditary nonpolyposis colorectal cancer gene carriers by screening for tumor microsatellite instability before germline genetic testing. *Gastroenterology* **120:**21–30, 2001.
120. Vasen, H. F., van Ballegooijen, M., Buskens, E., et al. A cost-effectiveness analysis of colorectal screening of hereditary nonpolyposis colorectal carcinoma gene carriers. *Cancer* **82:**1632–1637, 1998.
121. Soravia, C., Van Der Klift, H., Brundler, M. A., et al. Prostate cancer is part of the hereditary non-polyposis colorectal cancer (HNPCC) tumor spectrum. *Am. J. Med. Genet.* **121A:**159–162, 2003.
122. Syngal, S., Fox, E. A., Li, C., et al. Interpretation of genetic test results for hereditary nonpolyposis colorectal cancer: implications for clinical predisposition testing. *JAMA* **282:**247–253, 1999.

32 Molecular Genetic Applications to the Diagnosis of Lymphoma

William N. Rezuke Jr. and Gregory J. Tsongalis

1. INTRODUCTION

One of the earliest descriptions of primary tumors of lymph nodes was by Thomas Hodgkin in 1832. His publication entitled "On Some Morbid Appearances of the Absorbent Glands and Spleen" reported the clinical characteristics and post mortem findings from seven patients with an unusual disease of lymph nodes *(1)*. Some 30 yr later, the term "Hodgkin's disease" was introduced by Sir Samuel Wilks in honor of the work of Thomas Hodgkin *(2)*.

Today, the malignant lymphomas are regarded as a heterogeneous group of neoplasms, which can be broadly categorized into two groups: Hodgkin's lymphoma (formerly Hodgkin's disease) and non-Hodgkin's lymphoma. These neoplasms most often arise in lymphoid tissues such as lymph nodes and spleen, but could arise virtually anywhere in the body. Malignant lymphoma comprises approx 5% of new cancer cases in the United States annually, with approx 53,400 new cases of non-Hodgkin's lymphoma and 7,600 new cases of Hodgkin's lymphoma expected in 2003 *(3)*.

Precise diagnosis and classification of lymphoma is important because treatment options and prognosis vary considerably. The diagnosis of lymphoma prior to 1980 was based primarily on the histological evaluation of traditional hematoxylin and eosin (H&E) stained slides by light microscopy. In the 1980s, our understanding of the immune system and ability to diagnose and classify non-Hodgkin's and Hodgkin's lymphoma improved significantly, largely the result of development of immunopathological methods. These methods utilize a wide variety of commercially prepared monoclonal antibodies to detect cellular antigens. In the mid-1980s, the availability of molecular genetic methods further enhanced our ability to diagnose and classify lymphoid neoplasms.

The diagnosis and classification of lymphoma today requires the integration of traditional morphologic findings by light microscopy with a variety of special techniques, including immunopathological techniques (immunohistochemistry, flow cytometry), molecular genetic techniques (polymerase chain reaction [PCR], Southern blotting, and fluorescence *in situ* hybridization [FISH]), and cytogenetics. Of these, the routine application of immunopathological techniques has made the greatest contribution to our ability to diagnose and classify both Hodgkin's lymphoma and non-Hodgkin's lymphoma (Fig. 1A,B). Molecular genetic techniques are employed primarily in cases of non-Hodgkin lymphoma and only rarely in Hodgkin lymphoma. However, cases requiring molecular genetic applications are often the most diagnostically challenging cases for the pathologist. The objective of this chapter is to overview the hematopathological approach to the diagnosis of lymphoma, with an emphasis on the application of molecular genetic methods.

2. NON-HODGKIN'S LYMPHOMA

Non-Hodgkin's lymphoma comprises a heterogeneous group of lymphoid neoplasms, which occur because of neoplastic transformation of B- and T-lymphocytes at different stages of normal B-cell and T-cell development *(4,5)*. The wide variety of B-cell and T-cell non-Hodgkin's lymphomas reflects the varying stages of lymphocyte development and the complexity of the immune system. The clinical and pathological characteristics of the non-Hodgkin's lymphomas are summarized in a comprehensive manner in the recently published World Health Organization Classification of Tumours of Haematopoietic and Lymphoid Tissues (WHO Classification) *(6)*.

Centrally important to establishing a diagnosis of non-Hodgkin's lymphoma is the determination of clonality because both B- and T-cell non-Hodgkin's lymphomas represent monoclonal proliferations of B-cells and T-cells, respectively. A major application of molecular genetic methods in the evaluation of non-Hodgkin's lymphoma involves the determination of B- and T-cell clonality. These methods are considered to be the "gold standard" for determining clonality and are utilized primarily when clonality cannot be determined by immunopathological methods. For B-cell neoplasms, clonality can often be determined immunopathologically by demonstrating the presence of monoclonal surface immunoglobulin *(7)*. In contrast, for T-cell malignancies, there is no immunopathological

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

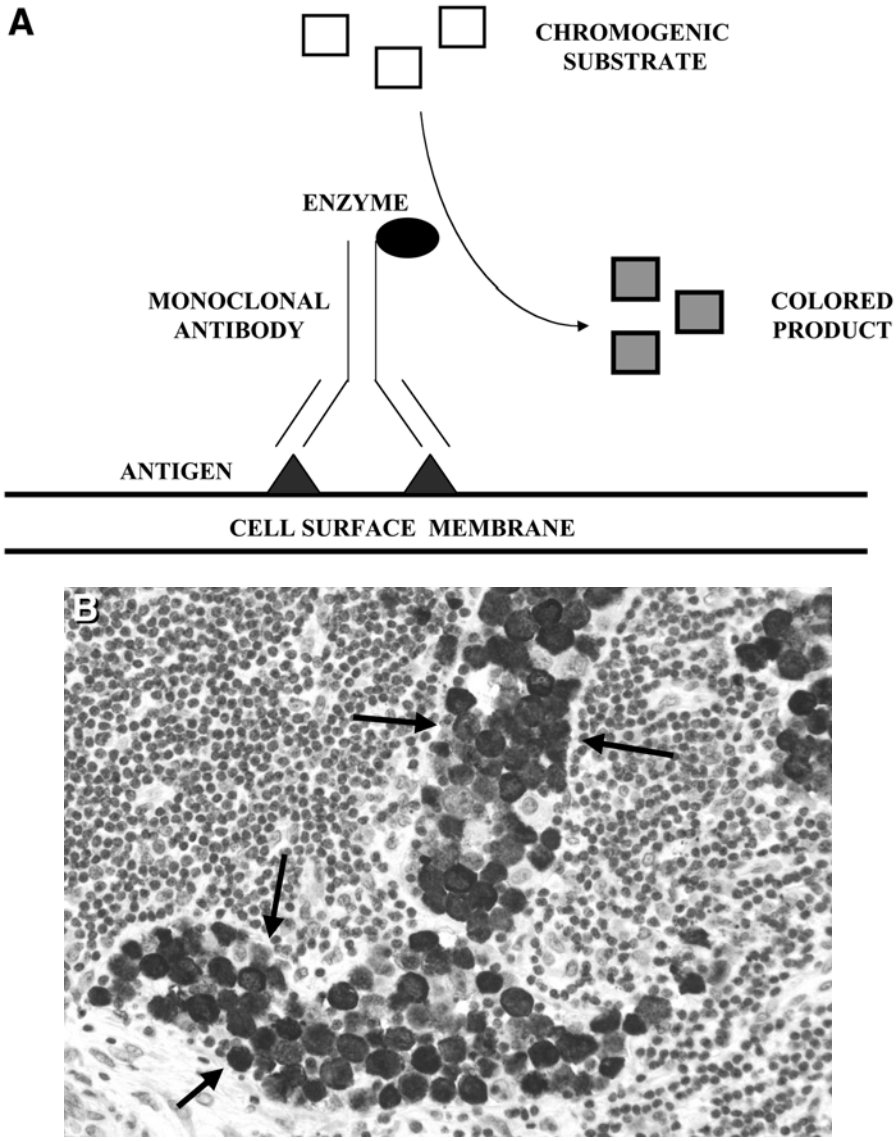


Fig. 1. Immunopathological detection of cell surface antigens. **(A)** Diagram illustrating the basic principle of immunopathological methods. An enzyme is linked to a monoclonal antibody, which recognizes a specific cell surface membrane antigen. A chromogenic substrate is added and is enzymatically converted to a colored product, which is visualized microscopically. **(B)** Lymph node biopsy with sinusoidal involvement by anaplastic large-cell lymphoma (ALCL). By immunoperoxidase, the neoplastic large lymphoid cells are strongly positive for ALK protein, characteristic of ALCL (arrows). The surrounding non-neoplastic small lymphocytes are negative.

equivalent to monoclonal surface immunoglobulin, although qualitative or quantitative loss of normal T-cell antigen expression is considered to be presumptive evidence of T-cell neoplasia *(7)*. Thus, molecular genetic approaches for the determination of clonality in T-cell non-Hodgkin's lymphoma are especially important.

Other molecular genetics applications to the assessment of lymphoid malignancies include determination of B-cell or T-cell lineage, detection of chromosomal translocations, and detection of minimal residual disease. The latter application has become increasingly useful in evaluating patients for minimal residual disease before and after bone marrow transplantation *(8)*. The detection of a specific chromosomal translocation could be very useful in the subclassification of non-Hodgkin's lymphoma. For example, in a lymph node with suspected follicular lymphoma, the detection of a translocation, t(14;18), involving the *BCL-2* proto-oncogene, would confirm this diagnosis. Similarly, the detection of a translocation, t(11;14), involving the *BCL-1* proto-oncogene, would confirm a diagnosis of mantle cell lymphoma *(9)*.

2.1. NORMAL B-CELL DEVELOPMENT According to current concepts of the normal humoral immune system, all B-lymphocytes arise from pluripotent stem cells in the bone marrow and subsequently migrate to secondary lymphoid organs such as lymph node follicles and Peyer's patches in the gastrointestinal (GI) tract. The stages of B-cell differentiation in the bone marrow occur largely independent of the presence of an antigen, whereas the stages of differentiation in secondary lymphoid organs require the presence of an antigen for transformation *(4,10)*.

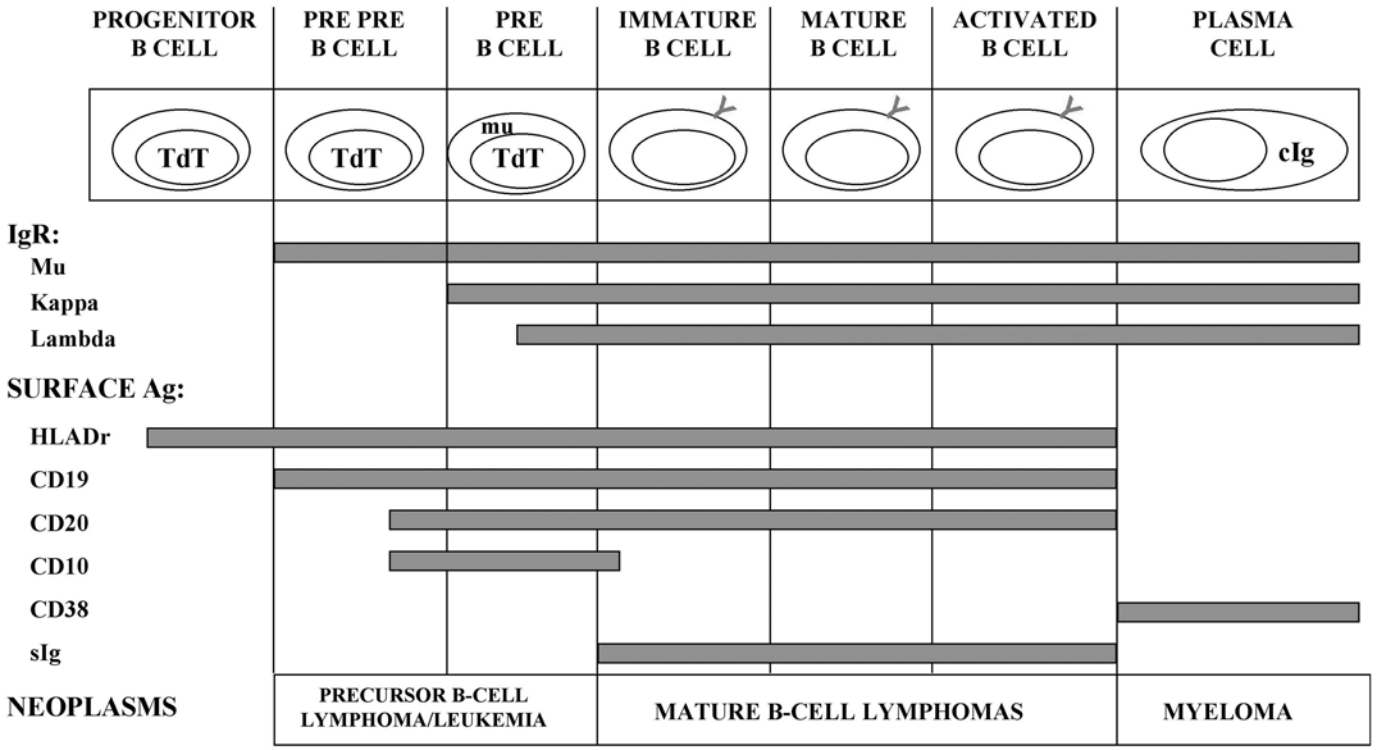


Fig. 2. Normal stages of B-cell development with immunological and molecular genetic changes at different stages. See text for discussion. TdT, terminal deoxynucleotidyl transferase; mu, cytoplasmic mu heavy chain; cIg, cytoplasmic immunoglobulin; IgR, immunoglobulin rearrangement; sIg, surface immunoglobulin.

Figure 2 shows the normal stages of B-cell development, which occur in an orderly fashion, beginning with a progenitor B-cell and maturing to a terminally differentiated plasma cell. A variety of recognized changes occur at different maturational stages both at the molecular level and with regard to the presence of specific cellular antigens. At the molecular level, the genes, which code for the immunoglobulin heavy- and light-chain proteins, undergo sequential rearrangements early in B-cell development (Fig. 2). Initially, the immunoglobulin mu heavy chain located on chromosome 14q32 undergoes rearrangement, followed by kappa light-chain rearrangement on chromosome 2p12 and lambda light-chain rearrangement on chromosome 22q11 *(11)*. Subsequent transcription and translation of the mu heavy-chain gene results in the appearance of cytoplasmic mu heavy-chain protein, which defines the pre-B-cell stage of development. The subsequent transcription and translation of kappa or lambda light-chain genes results in an intact surface immunoglobulin receptor, which consists of two heavy chains and two light chains (Fig. 3A). An intact surface immunoglobulin receptor is characteristically present in the immature, mature, and activated B-cell stages.

As illustrated in Fig. 2, a variety of cellular antigens can be detected at different stages in normal B-cell development. CD (cluster designation) numbers have been applied to many of these antigens to simplify terminology. The earliest antigens identified on B-cells are terminal deoxynucleotydil transferase (TdT) and HLA-Dr. Neither of these antigens are B-lineage-specific. Specific B-lineage-associated antigens, CD19 and CD20, as well as the common ALL antigen, CD10, are subsequently expressed by B-cells. As a B-cell matures to a terminally differentiated plasma cell, the majority of B-cell surface antigens are no longer expressed and the CD38 antigen appears. The plasma cell is also characterized by presence of cytoplasmic immunoglobulin.

The fundamental theory of lymphoid neoplasia is that disorders of lymphoid cells represent neoplastic transformation of cells at various stages of normal development *(12)*. For example, precursor B-cell acute lymphoblastic lymphoma/leukemia immunophenotypically mimics normal precursor B-cells by showing expression for TdT, HLA-Dr, CD10, CD19, variable CD20, and (cytoplasmic) mu heavy chain (Fig. 2). Examples of non-Hodgkin's lymphoma, which are neoplastic counterparts to relatively mature B-lymphocytes, include chronic lymphocytic leukemia/small lymphocytic lymphoma, follicular lymphoma, and mantle cell lymphoma. Plasma cell myeloma is the neoplastic counterpart of the normal plasma cell.

2.2. NORMAL T-CELL DEVELOPMENT T-Lymphocytes also arise from pluripotent stem cells in the bone marrow. In contrast to B-cell development in which the earliest stages of maturation occur in the bone marrow, progenitor T-cells migrate from the bone marrow to the thymus, where the early stages of T-cell development occur *(5,13)*. Subsequently, mature T-cells circulate in the peripheral blood and seed peripheral lymphoid tissues, which include paracortical areas of lymph nodes and periarteriolar sheaths of the spleen.

Figure 4 shows the normal stages of T-cell development in the thymus, which, analogous to B-cell development, occur in an orderly fashion. T-Lymphocytes possess a surface membrane protein complex referred to as the T-cell receptor, which is structurally similar to the B-cell surface immunoglobulin

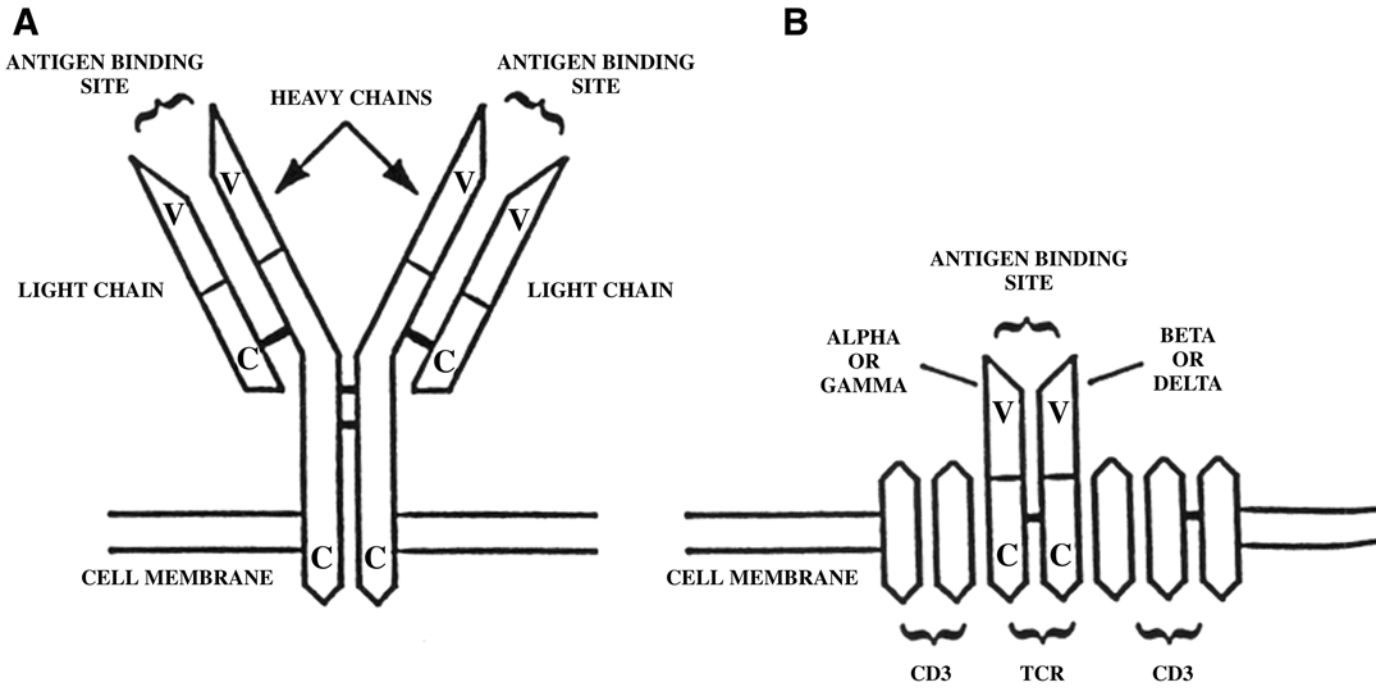


Fig. 3. Schematic diagram of immunoglobulin and T-cell receptors. (**A**) The immunoglobulin protein is a heterodimer composed of two heavy and two light chains, each of which has variable (V) and constant (C) regions. (**B**) The T-cell receptor (TCR) is also a heterodimer composed of either one α (alpha)- and one β (beta)-chain or one γ (gamma)- and one δ (delta)-chain. Each of the TCR proteins has variable (V) and constant (C) regions. CD3 is a complex of five proteins associated with the TCR.

receptor (Fig. 3B). The T-cell receptor is a heterodimer protein comprised of either one α (alpha) and one β (beta) polypeptide chain, which combine to form the αβ T-cell receptor, or one γ (gamma) and one δ (delta) polypeptide chain, which combine to form the γδ T-cell receptor *(14–16)*. Approximately 98–99% of circulating T-cells possess the αβ T-cell receptor and the remaining 1–2% of T-cells possess the γδ T-cell receptor *(16)*. The genes that code for each of the polypeptide chains that comprise the T-cell receptor undergo sequential rearrangements early in T-cell development. The first T-cell receptor gene to rearrange is δ, which is followed sequentially by γ, β, and α genes. The genes coding for the α and δ polypeptide chains are located on chromosome 14q11, the β-chain gene on chromosome 7q34, and the γ-chain gene on chromosome 7p15 *(11)*.

Analogous to developing B-cells, a variety of cellular antigens can be detected at different stages of normal T-cell development (Fig. 4). The earliest antigens expressed are TdT and the T-lineage antigen CD7. The CD3 antigen, which is a protein complex associated with the T-cell receptor (Fig. 3B), is primarily a cytoplasmic antigen early in T-cell development and is expressed as a cell surface membrane antigen at later stages. The common thymocyte stage of T-cell development is defined by expression of CD1a, the common thymocyte antigen, and is characterized by coexpression of the CD4 (helper/inducer) and CD8 (cytotoxic/suppressor) T-cell subset antigens. Phenotypically mature T-cells (also referred to as peripheral T-cells) express either the CD4 antigen or the CD8 antigen, but not both. Similar to B-cell non-Hodgkin's lymphoma, the T-cell neoplasms represent neoplastic transformation of cells at various stages of normal T-cell development *(12)*. For example, precursor T-cell lymphoblastic lymphoma/leukemia might exhibit an immunophenotype, which resembles the normal common thymocyte, showing expression for TdT, CD1a, and cytoplasmic CD3 and CD7, as well as coexpression for CD4 and CD8 (Fig. 4). Examples of T-cell non-Hodgkin's lymphoma, which are neoplastic counterparts to relatively mature or peripheral T-lymphocytes, include anaplastic large-cell lymphoma, cutaneous T-cell lymphoma (mycosis fungoides), and peripheral T-cell lymphoma *(6)*.

2.3. B-CELL SURFACE IMMUNOGLOBULIN AND T-CELL RECEPTOR GENE REARRANGEMENT The B-cell surface immunoglobulin receptor and T-cell receptor are involved in the process of antigen recognition by normal B-cells and T-cells, respectively. These receptors are structurally similar, being heterodimer proteins linked by disulfide bonds, and are composed of both variable (V) and constant (C) regions *(11)* (Fig. 3). The variable regions of these proteins are similarly involved in antigen recognition. The constant region of the immunoglobulin heavy-chain protein defines the different immunoglobulin subclasses (IgG1, IgG2, IgG3, IgG4, IgA1, IgA2, IgM, IgD, and IgE) *(10)*. The genes that code for the B- and T-cell receptors are also structurally similar and consist of a large number of exons or coding sequences referred to as a *supergene family*. The exons undergo a process of DNA recombination or rearrangement leading eventually to transcription and translation and the production of functional receptor proteins *(10,11,14,16,17)*.

A general scheme illustrating the process of B-cell surface immunoglobulin and T-cell receptor gene rearrangement is shown in Fig. 5. The *germline configuration* refers to nonrearranged

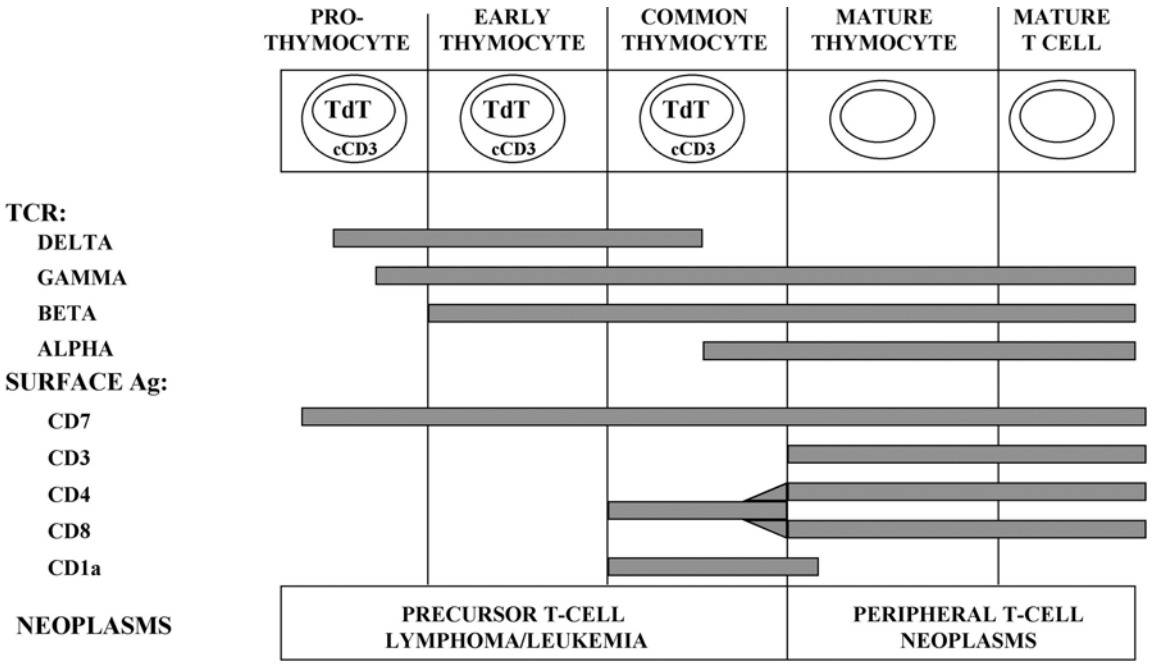


Fig. 4. Normal stages of T-cell development with immunological and molecular genetic changes at different stages. See text for discussion. TdT, terminal deoxynucleotidyl transferase; cCD3, cytoplasmic CD3; TCR, T-cell-receptor rearrangements; Ag, antigen.

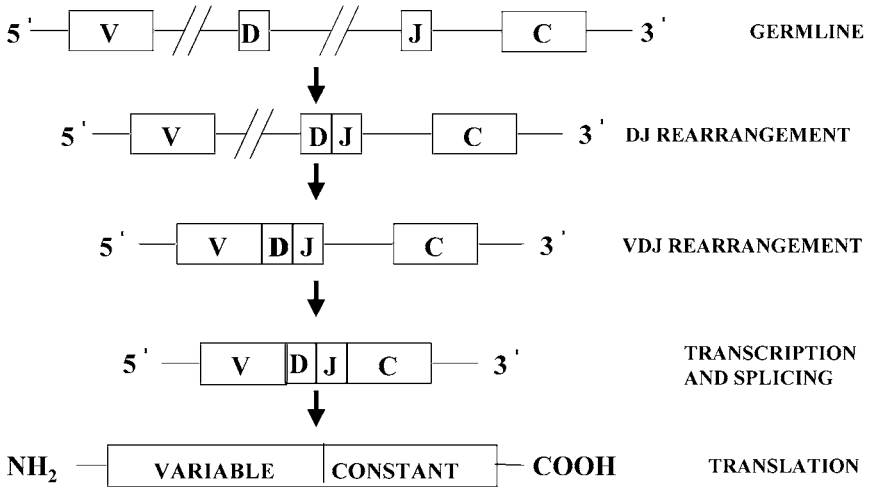


Fig. 5. Schematic diagram illustrating the sequential steps involved in immunoglobulin and T-cell receptor gene rearrangements. See text for discussion. V, variable segments; D, diversity segments; J, junctional segments; C, constant segments.

DNA. The exons that code for the variable regions of the immunoglobulin and T-cell receptors are referred to as variable (V) segments, diversity (D) segments, and junctional (J) segments, and those that code for the constant regions are referred to as (C) segments. The process of gene rearrangement first involves the selective apposition of one D segment with one J segment by deletion of the intervening coding and noncoding DNA sequences, resulting in a DJ rearrangement. A similar process of rearrangement apposes a V segment, located in the 5′ direction to D and J, to form a VDJ rearrangement. Transcription to messenger RNA (mRNA) then occurs even though the rearranged VDJ segments are not yet directly apposed to C segments, which are remotely located in the 3′ direction. Subsequent splicing of the mRNA with deletion of noncoding sequences results in apposition of VDJ with C to form a VDJC mRNA, which can then be translated into a surface immunoglobulin or T-cell-receptor protein.

The genes coding for the immunoglobulin heavy-chain proteins and T-cell-receptor β- and δ-chain proteins include V, D, J, and C segments. The genes coding for the kappa and lambda light-chain proteins and the T-cell-receptor α and γ proteins include only V, J, and C segments without D segments *(11,14,17)*. A schematic diagram of the immunoglobulin heavy chain and the T-cell-receptor β-chain supergene families is shown in Fig. 6. The immunoglobulin heavy-chain gene consists of more than 100 V_H segments, approx 30 D segments, 9

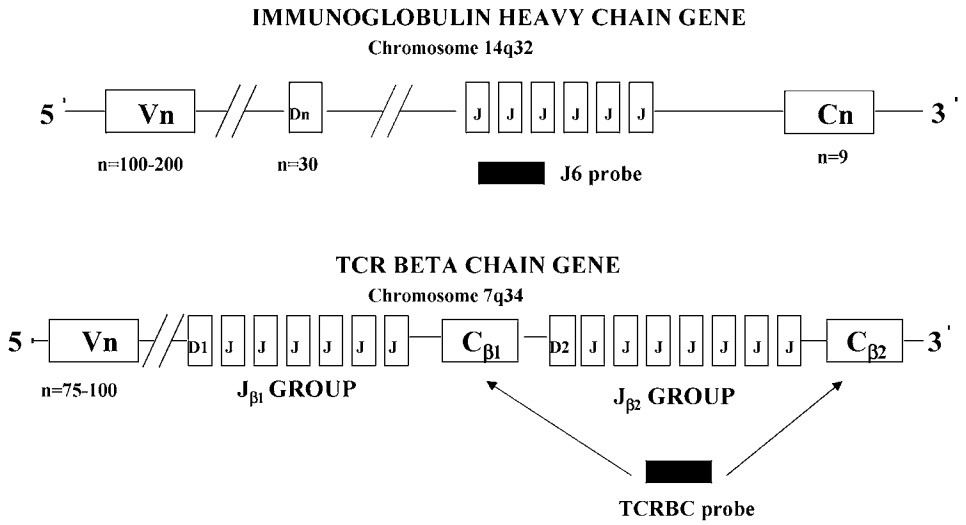


Fig. 6. Schematic diagram of the immunoglobulin heavy chain and the T-cell-receptor β-chain supergene families. The specific sites of recognition for probes used for Southern blotting are typically associated with consensus sequences in the joining (J_6 probe) and constant regions (TCRBC probe).

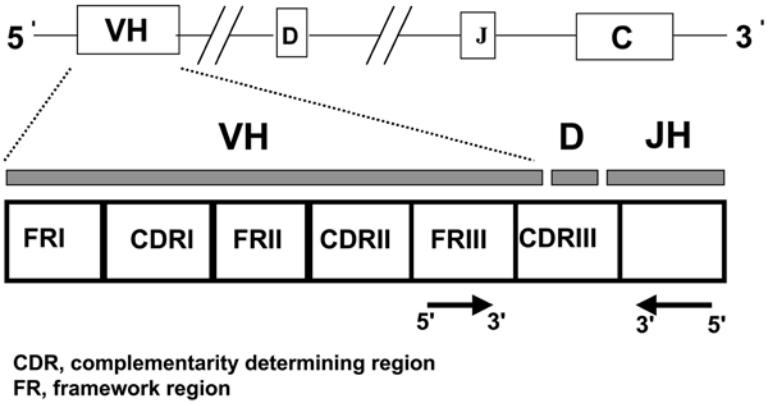


Fig. 7. An individual immunoglobulin V_H segment consists of framework regions (FRs), which are conserved nucleotide sequences, and complementarity regions (CDRs), which are hypervariable DNA sequences that code for the antigen-binding site and tend to undergo somatic hypermutation. Primers sites for PCR amplification are also shown (arrows). Primer sets for PCR analysis include one consensus V_H (FRIII region) primer and one consensus J_H primer.

J_H segments, and 11 C_H segments. The V_H segments can be further divided into seven families based on the presence of conserved or homologous DNA sequences, which are common to each member of a family. The V_H families are designated V_H1 to V_H7. The T-cell-receptor β-chain gene includes 75–100 V_β segments, which can be further divided into several families, and two tandem DJC complexes referred to as D1J1C1 and D2J2C2. Each DJC complex contains one D segment and one C segment. The first DJC complex contains six J_β segments ($J_{\beta1}$ group) and the second DJC complex contains seven J_β segments ($J_{\beta2}$ group) *(11,14,17)*.

The humoral (B-cell) and cell-mediated (T-cell) systems must be able to recognize a wide variety of environmental antigens even though only a finite quantity of DNA is present in each cell. The complex process of DNA recombination or rearrangement involving the surface immunoglobulin and T-cell receptor allows for tremendous diversity of both the humoral and cell-mediated systems and, hence, the ability of these systems to detect a wide variety of antigens *(10,11,14,16,17)*. The large number of V, D, J, and C segments results in many combinations, which can be transcribed and translated to millions of different antigen receptors.

Further diversification of the immunoglobulin receptor genes is achieved through a process of *somatic hypermutation*, which occurs normally within germinal center B-cells. In addition to conserved or homologous DNA sequences within V segments referred to as *framework regions* (FR), individual immunoglobulin heavy- and light-chain V segments contain *complementarity-determining regions* (CDR), which contain nucleotide sequences that encode for the antigen-binding site (Fig. 7). The CDRs contain hypervariable nucleotide sequences, which tend to undergo somatic hypermutation. This process results in a series of DNA point mutations with consequent amino acid substitutions within the antigen-binding site. The end product of somatic hypermutation involving the CDRs is enhanced antibody affinity and specificity for target antigens *(10,11,14,16,17)*.

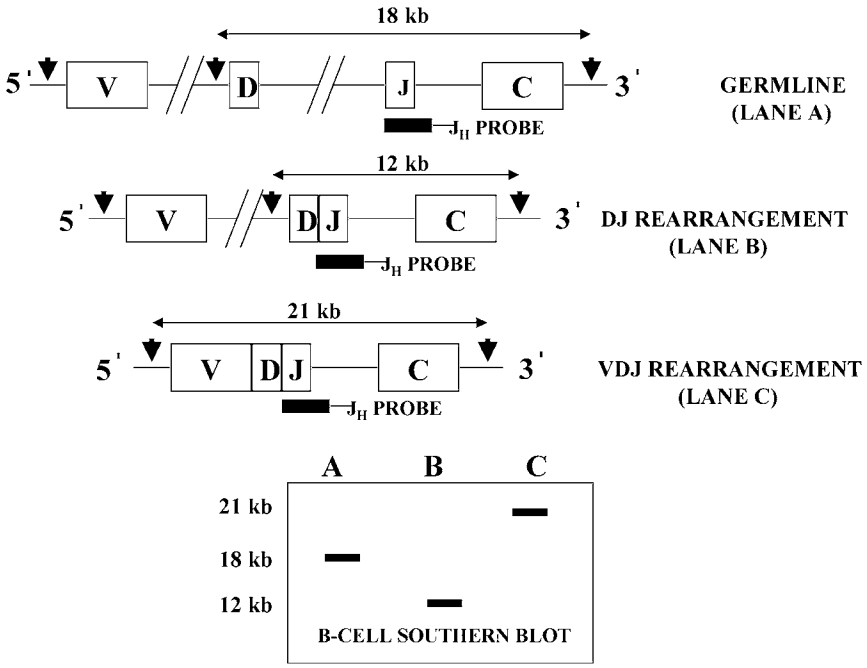


Fig. 8. Schematic diagram illustrating the Southern blot approach for detecting a B-cell heavy-chain gene rearrangement. Arrowheads identify restriction enzyme cleavage sites. See text for discussion.

3. THE SOUTHERN BLOT AND THE DETERMINATION OF B- AND T-CELL CLONALITY

To establish a diagnosis of B-cell or T-cell non-Hodgkin's lymphoma, the ability to prove that a neoplastic population of B-cells or T-cells is monoclonal in origin is of central importance. A monoclonal, or simply, a clonal cell population refers to a population of cells, which are all derived from a single precursor cell and thus share similar characteristics.

In lymphoid malignancies, clonality can be defined in several different ways. By light microscopy, clonality can be inferred based on the morphologic finding of a uniform or monomorphous population of lymphocytes. By immunopathological methods, a monoclonal B-cell population is defined by the presence of monoclonal surface immunoglobulin; that is, a population of B-cells uniformly expresses either a kappa light chain or a lambda light chain. In contrast, a benign or reactive B-cell population is characteristically polyclonal; that is, the B-cells include a mixture of kappa light chain- and lambda light chain-bearing cells. For T-cell neoplasms, there is no specific immuopathological marker of clonality, although qualitative change or aberrant loss of T-cell antigen expression could be considered presumptive evidence of T-cell clonality. By cytogenetics, a clonal population is defined as a population of cells with a recurrent chromosomal alteration such as recurrent translocation. Finally, by molecular genetic methods, clonality is defined by the identification of a clonal B-cell or T-cell gene rearrangement. The evaluation of clonality in any suspected case of non-Hodgkin's lymphoma first employs immunopathological methods and then molecular genetic methods if immunopathological studies are inconclusive.

The Southern blot is a very sensitive and specific method for determining B-cell and T-cell clonality and might detect a monoclonal population, which comprises as little as 1–5% of the total cell population *(11,14,18)*. Because the Southern blot requires high-quality DNA, fresh or frozen specimens are most suitable for analysis. Optimal specimens include cell suspensions or frozen/cryostat sections prepared from tissues such as lymph node, spleen, or cell suspensions prepared from peripheral blood specimens, bone marrow aspirates, body fluids, and fine-needle aspirates. For Southern blot analysis, DNA is first extracted and purified from the cell suspension or tissue sample. Separate samples of purified DNA are then digested with three different restriction enzymes suitable for the probe being used *(19–21)*. Restriction enzymes cleave DNA at specific sites by recognizing specific basepair sequences. The digested DNA fragments are then electrophoresed using agarose gels, which separate the DNA fragments according to molecular size. These DNA fragments are then transferred to a nylon membrane and hybridized with a specific DNA probe. DNA probe detection systems include radioactive labeling with ^{32}P, chemiluminescence, and colorimetry *(20–22)*.

Examples of DNA probes used to evaluate for the presence of a monoclonal B-cell population include a J_6 probe, which recognizes the immunoglobulin heavy chain joining (J) segments, and a J_K probe, which recognizes the immunoglobulin kappa light chain joining (J) segments. An example of a DNA probe used to evaluate for the presence of a monoclonal T-cell population is the TCRBC probe, which recognizes the two β-chain constant (C) segments. Figure 6 illustrates the specific sites of recognition for the J_6 probe in the immunoglobulin heavy-chain gene complex and the TCRBC probe in the T-cell-receptor β-chain gene complex.

The Southern blot approach for the detection of a clonal B-cell heavy-chain gene rearrangement is shown schematically in Fig. 8. In reactive or polyclonal B-lymphocyte populations, the primary band identified with the J_H probe is the germline band (lane A). Thousands of different VDJ rearrangements are

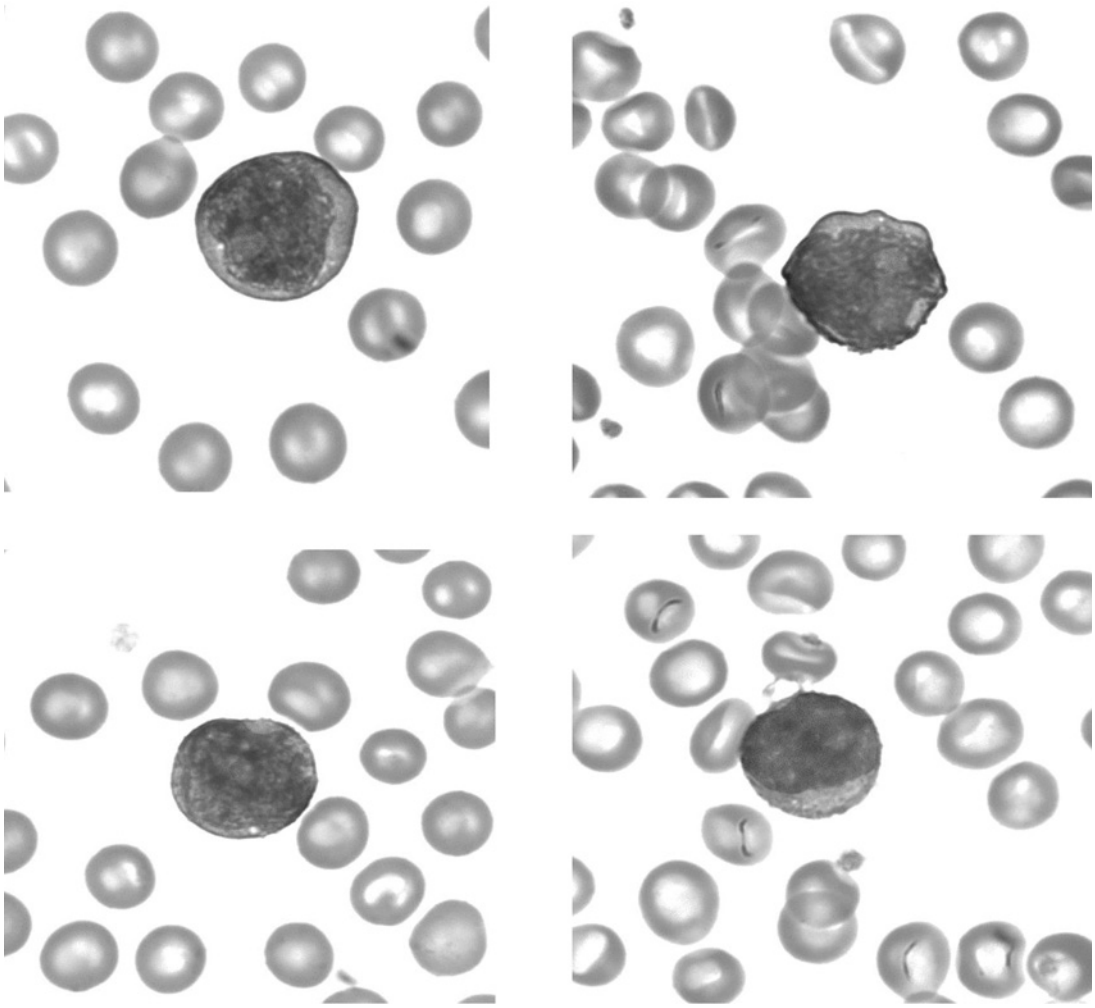


Fig. 9. Peripheral blood smear showing the presence of a uniform population of cytologically abnormal, small to medium-sized lymphocytes as described in the text.

actually present in this lane along with the germline band; however, the individual rearrangements are too small to be detected. In a monoclonal B-cell population, all B-cells are derived from a single precursor cell. Thus, all cells have an identical gene rearrangement, which will be detected by the Southern blot as a *novel band*. If the monoclonal B-cell population has a DJ rearrangement, numerous intervening coding and noncoding DNA sequences are deleted, resulting in a smaller fragment of DNA detected by the J_H probe (lane B). If the monoclonal B-cell population has a VDJ rearrangement, a restriction enzyme cleavage site is also deleted, resulting in a larger fragment of DNA detected by the J_H probe (lane C).

The presence of a clonal B- or T-cell gene rearrangement is established by Southern blot methodology by identifying at least two novel bands in any of the three enzyme digests. The novel bands can be present in the same enzyme digest or in separate enzyme digests *(19)*. A novel band refers to any band occurring in a lane other than the following:

1. A germline band
2. A cross-hybridization band, which occurs as a result of hybridization of the probe to partially homologous DNA sequences in other areas of the genome
3. A partial digest band, which occurs as a result of incomplete digestion of DNA by a restriction enzyme

3.1. CASE EXAMPLE INVOLVING APPLICATION OF THE SOUTHERN BLOT A 66-yr-old asymptomatic patient was referred to an oncologist because of an abnormal complete blood count (CBC), which showed the presence of a mild absolute lymphocytosis of 7500/mm^3 (normal range: 1500–4500/mm^3). Review of the patient's peripheral blood smear showed the presence of a uniform population of cytologically abnormal, small to medium-sized lymphocytes with generally round to slightly irregular nuclei, condensed chromatin, often single prominent nucleoli, and scant to moderate basophilic cytoplasm (Fig. 9). Flow cytometry studies were performed and demonstrated a predominant T-cell population, which was phenotypically abnormal. The T-cells were positive for the pan-T-cell antigens, CD2, CD3, CD5, and CD7; however, CD2 expression was qualitatively dimmer than normal and CD5 expression was qualitatively brighter than normal. The phenotypically abnormal T-cell population was also uniformly positive for the T-cell subset antigen CD4. Considered together, the morphologic and immunophenotypic findings were suspicious for the presence of a T-cell lymphoproliferative disorder; a

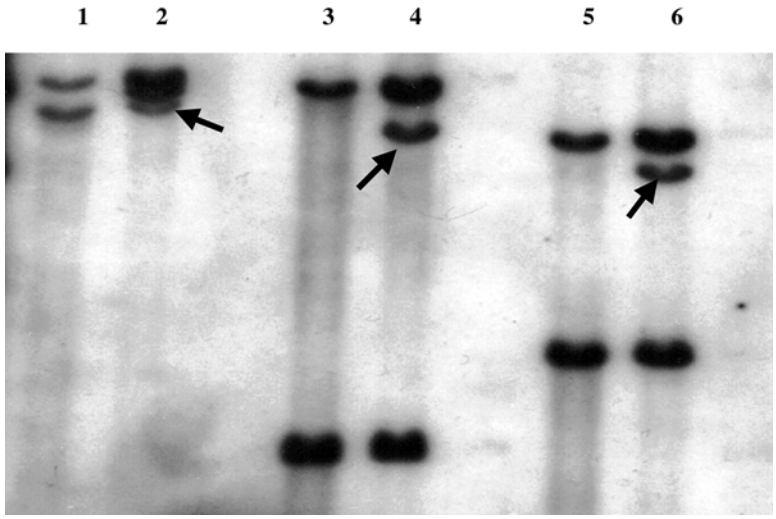


Fig. 10. Evaluation for evidence of T-cell clonality by Southern blot using the TCRBC probe (DAKO Corp., Carpenteria, CA) and restriction enzymes *Bgl* II (lanes 1 and 2), *Xba* 1 (lanes 3 and 4), and *Bam*HI/*Hin*dIII (lanes 5 and 6). Each restriction enzyme has a control lane identifying the germline configuration (lanes 1, 3, and 5). A clonal T-cell gene rearrangement is identified in the patient's sample (lanes 2, 4, and 6) based on the presence of a single novel band in each enzyme digest (arrows).

definitive diagnosis could not be established in this case in the absence of a definitive marker of T-cell clonality.

Molecular genetic studies to evaluate for T-cell clonality were subsequently performed. Analysis of the peripheral blood sample was initially performed by PCR using primers directed to the T-cell-receptor gamma gene region and failed to show evidence of a monoclonal band. The Southern blot result using a T-cell (TCRBC) probe chemiluminescence detection system and three sets of restriction enzymes is shown in Fig. 10. The Southern blot includes a marker lane that consists of predigested fragments of lambda phage DNA to assess restriction fragment sizes. With each enzyme digest, a control lane consisting of normal placental DNA is run to identify the germline configuration for each enzyme digest (lanes 1, 3, and 5). The patient's sample shows the presence a single novel band in each of the three enzyme digests (arrows, lanes 2, 4, and 6). The finding of at least two novel bands in this case is indicative of a clonal T-cell gene rearrangement. The Southern blot finding confirmed the diagnosis of a T-cell lymphoproliferative disorder. Subsequent cytogenetic studies demonstrated the presence of a t(14;14) (q11;q32), which is characteristically associated with T-cell prolymphocytic leukemia.

4. THE POLYMERASE CHAIN REACTION (PCR) AND THE DETERMINATION OF B- AND T-CELL CLONALITY

The PCR technique has become a popular method for evaluating for the presence or absence of B- and T-cell clonality in lymphoid neoplasms *(23–40)*. This powerful methodology for DNA analysis allows for the evaluation of minute quantities of DNA by in vitro amplification. Analogous to Southern blot methods, the application of PCR to detect B- and T-cell clonality involves evaluation of gene rearrangements in those segments of DNA, which code for the variable regions of the surface immunoglobulin and T-cell receptors. Each variable (V) segment of DNA has unique DNA sequences, referred to as complementarity regions, which contribute to the great diversity of the immunoglobulin and T-cell-receptor antigen recognition sites. In addition, short sequences of DNA are shared by nearly all of the V segments, which are referred to as framework regions (Fig. 7). The shared segments of each V segment can be recognized by a primer referred to as a *consensus* V region primer. In a similar fashion, short sequences of DNA shared by nearly all of the J segments can be recognized by a *consensus* J region primer *(26–43)*.

A diagram illustrating the application of PCR for the detection of a B-cell heavy-chain gene rearrangement using V_H and J_H consensus primers is shown in Fig. 11, and an ethidium bromide stained PCR gel is shown in Fig. 12. In order to successfully amplify a segment of DNA by PCR, the primers must recognize DNA sequences within a short segment of DNA. In the germline configuration, because V and J segments are widely separated, no significant DNA product is obtained following amplification by PCR (Fig. 11, lane A; and Fig. 12, lane 2). If a VDJ rearrangement occurs, the proximity of the V and J segments allows for successful amplification of a DNA product. Polyclonal B- or T-cell populations have large numbers of rearrangements, which differ slightly in size, resulting in a polyclonal smear or ladder pattern of DNA amplification (Fig. 11, lane B; Fig. 12, lanes 4 and 5). In contrast, monoclonal B- or T-cell cell populations contain an identical rearrangement, which results in the formation of a distinct band (Fig. 11, lane C; Fig. 12, lanes 6 and 7). The determination of an immunoglobulin heavy-chain (IgH) rearrangement by PCR involves one consensus primer directed to a V_H segment framework region (FRI, FRII, or FRIII) and one consensus J_H primer (Fig. 7).

Although the Southern blot method is considered to be the "gold standard" for demonstrating clonality in lymphoid neoplasms, PCR offers distinct advantages *(23,41–43)* (Table 1). Southern blotting is costly and labor-intensive requiring 7–10 d to obtain a result; PCR can be performed at a lower cost in just 1–2 d. The Southern blot method also requires a relatively large amount (at least 30 μg) of high-quality intact DNA, which, in

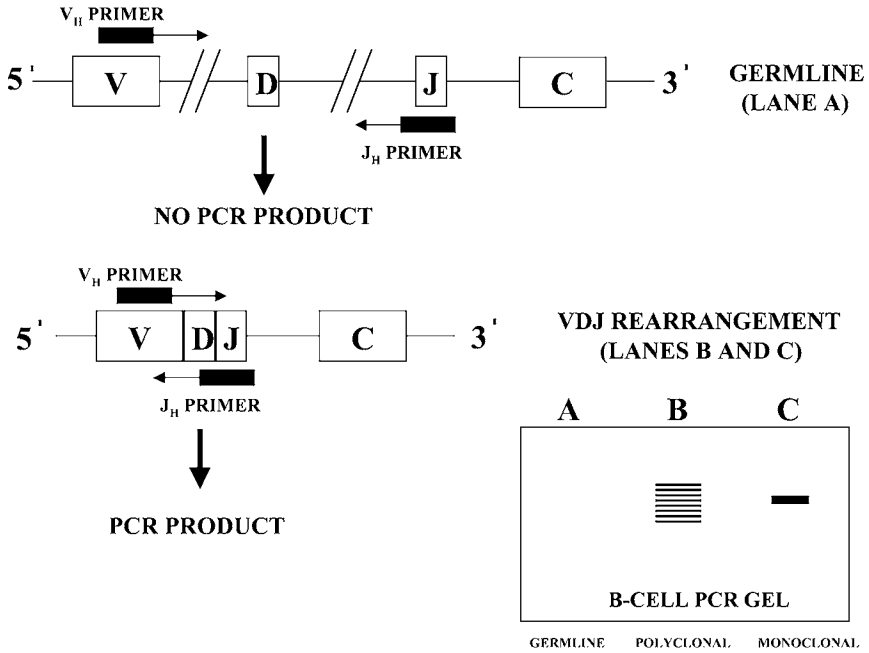


Fig. 11. Schematic diagram illustrating the PCR approach for detecting a B-cell gene rearrangement. See text for discussion.

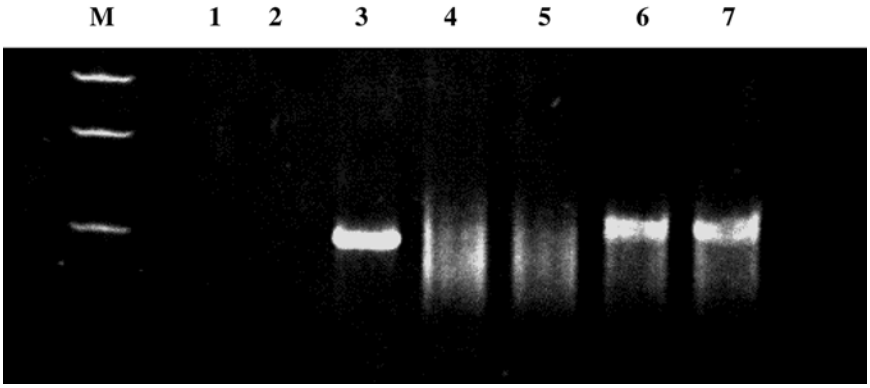


Fig. 12. B-Cell gene rearrangement patterns using J_H and V_H consensus primers. PCR gels are shown stained with ethidium bromide. In the germline configuration, no PCR product is obtained (lane 2). Polyclonal B-cell populations have a characteristic polyclonal smear or ladder pattern (lanes 4 and 5). Monoclonal B-cell populations are characterized by a single distinct band (lanes 6 and 7). Lane M, molecular size marker; lane 1, blank; lane 3, positive control.

Table 1
Comparison of Southern Blot and PCR Technologies

	Southern blot	*PCR*
Labor	More	Less
Cost	More	Less
Sensitivity	1–5% clone	0.1% clone
DNA quantity	Large	Small
DNA quality	High	High/low
Tissues	Fresh	Fresh/paraffin
False Negatives	Low	High

most cases, can only be obtained from fresh or frozen tissue samples. In contrast, the amplification of DNA by PCR requires only short segments of DNA, allowing this type of analysis to be performed on small concentrations of DNA and on DNA that is of low quality or only partially intact (such as DNA extracted from paraffin-embedded tissues) *(44–46)*. Finally, although Southern blotting might detect a 1-5% clonal lymphoid population, PCR might detect as small as a 0.1% clonal lymphoid population *(25)*.

Despite the many advantages of PCR in evaluating for B- and T-cell clonality, the technique is associated with a higher percentage of false-negative results compared to Southern blotting *(44,45)*. This high false-negative rate likely occurs because of the inability of consensus V primers to recognize complementary DNA sequences in all of the V segments and because of the inability of V and J primers to recognize genetic alternations such as partial rearrangements (DJ rearrangements), chromosomal translocations, and somatic hypermutations involving the antigen receptor gene loci *(23,47)*.

4.1. CASE EXAMPLE OF APPLICATION OF PCR A 68-yr-old patient presented with a history of gastrointestinal bleeding. Endoscopic examination of the stomach revealed thickened gastric folds with multiple small ulcers. Microscopic

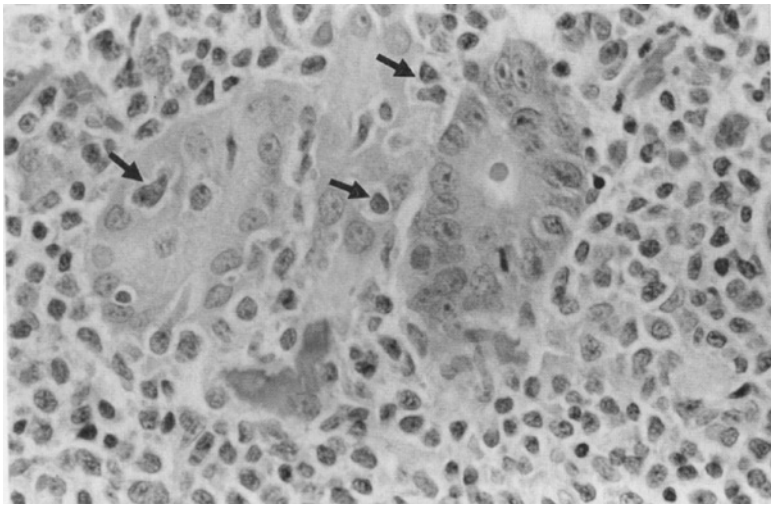


Fig. 13. Histological section of the gastric biopsy showing infiltration of normal gastric glands by abnormal small lymphoid cells (arrows) forming lymphoepithelial lesions. H & E stain; magnification: ×500.

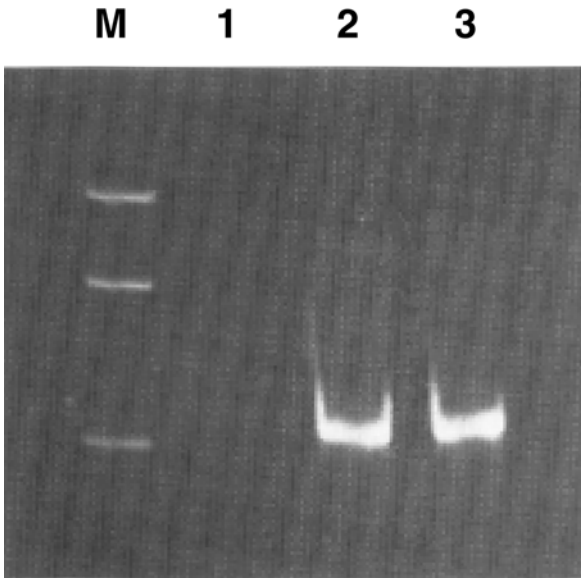


Fig. 14. PCR gel stained with ethidium bromide after amplification of DNA with $J_H$and V_H consensus primers. A single distinct band is present indicative of a monoclonal B-cell population. Studies performed in duplicate, lanes 2 and 3. Lane M, molecular size marker; lane 1, blank.

examination of the endoscopic biopsies obtained adjacent to the gastric ulcers showed a dense lymphoid infiltrate with a varied histological appearance. The infiltrate was characterized by the presence of obviously benign, reactive germinal centers, as well as sheets of cytologically atypical small lymphoid cells. The small lymphoid cells focally infiltrated normal gastric glands forming so-called lymphoepithelial lesions (Fig. 13). In areas, gastric glands were obliterated.

The microscopic findings suggested a differential diagnosis, which included chronic gastritis and a relatively common lymphoma in the stomach referred to as an extranodal marginal zone B-cell lymphoma (or low-grade B-cell lymphoma of mucosa-associated lymphoid tissue). To distinguish chronic gastritis from a B-cell non-Hodgkin's lymphoma (NHL), it was necessary to evaluate the biopsy for evidence of clonality; B-cell NHL would be expected to be a monoclonal process. By immunopathology, the sheets of atypical small lymphoid cells were diffusely positive for the B-cell antigen CD20; however, kappa and lambda light-chain studies were inconclusive. Thus, by immunopathologic studies, clonality was indeterminate. Molecular genetic analysis by PCR was then performed to further evaluate the biopsy for evidence of B-cell clonality. PCR studies were performed using J_H and V_H (FR III) consensus primers directed to the B-cell heavy-chain gene region and showed the presence of a single distinct band, indicative of a monoclonal B-cell population (Fig. 14).

A diagnosis of extranodal marginal zone B-cell lymphoma was established in this case based on the demonstration of a monoclonal B-cell population by PCR in combination with the characteristic histologic and immunopathologic findings. These lymphomas frequently arise in lymphoid tissue associated with mucosal sites, such as the gastrointestinal tract, lung, and salivary gland and are often difficult to distinguish histologically from chronic inflammatory disorders such as chronic gastritis. This case illustrates a common problem confronting the pathologist and demonstrates the contribution of PCR to resolving these problems.

5. FLUORESCENT *IN SITU* HYBRIDIZATION

Molecular genetic analysis of NHL by FISH has become an increasingly popular method of detecting chromosomal abnormalities resulting from the relative simplicity of the procedure and the commercial availability of a large number of chromosome specific probes. The technique of FISH is based on the ability of single-stranded DNA sequences (target DNA) to anneal to complementary DNA sequences (fluorescent DNA probe). The target DNA sequences might be present in interphase cells or metaphase cells, which are affixed to a glass slide. After incubation of the target DNA with a fluorescent DNA probe, the result is interpreted microscopically. FISH can be performed on peripheral blood, bone marrow aspirate, and body fluid smears, as well as on histological sections *(48–50)*.

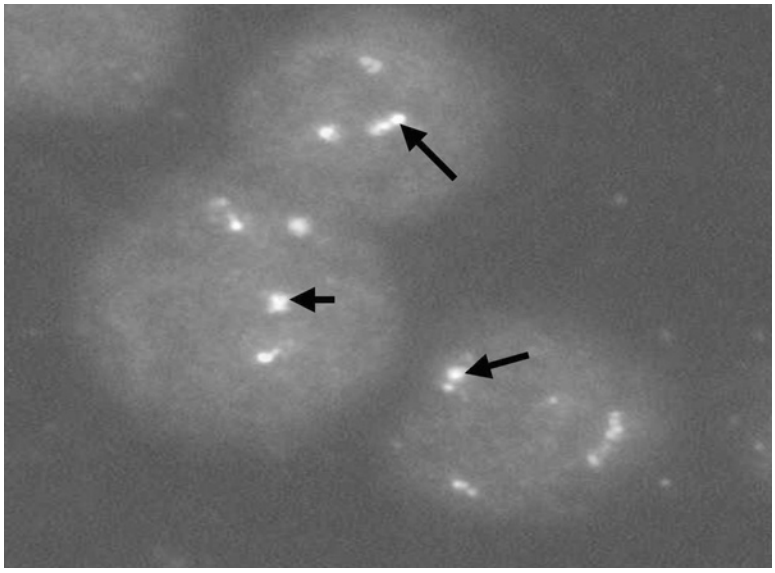


Fig. 15. FISH analysis utilizing a dual-color translocation probe system, which includes a probe directed to the immunoglobulin heavy-chain gene region (14q32) labeled with SpectrumGreen (green signal) and a probe directed to the *BCL-1/CCND1* gene region (11q13) labeled with SpectrumOrange (orange signal). A yellow signal (arrow) occurs because of fusion of the red and green signals and indicates the presence of a t(11;14)(q13;q32). (Figure appears in color in insert following p. 172.)

The NHLs are especially suited for analysis by FISH because of the presence of a number of recurrent, nonrandom chromosomal aberrations, which can be broadly grouped into (1) molecular translocations, (2) numerical chromosomal aberrations, and (3) gene deletions *(48)*. The most frequent application of FISH in the setting of NHL involves evaluation for evidence of a translocation, such as the t(11;14) (q13;q32) associated with mantle cell lymphoma. *Translocation probes* are used in this setting; typically, these are very long probes spanning long segments of DNA, including the DNA breakpoints associated with the translocation. Numerical chromosomal aberrations refer to loss or gain of a whole chromosome and can be detected using *centromeric probes*. Centromeric probes hybridize to chromosome-specific, repetitive DNA sequences located near the centromeres. Evaluation for evidence of a trisomy 3 associated with marginal zone B-cell lymphoma and trisomy 12 associated with chronic lymphocytic leukemia/small lymphocytic lymphoma would involve the use of centromeric probes. Finally, evaluation for evidence of a gene deletion, such as deletion of a tumor suppressor gene, would involve *gene-specific probes (48,49)*.

Analysis for evidence of chromosomal abnormalities by FISH in the evaluation of NHL offers several advantages over traditional karyotypic analysis by cytogenetics. Cytogenetic studies are especially problematic in the evaluation of low-grade or indolent NHL, which is characteristically associated with a low cell proliferation rate. Cytogenetic analysis requires metaphase spreads and, therefore, viable and proliferating cells; thus, cytogenetic analysis of low-grade NHL is frequently unsuccessful. The low mitotic rate of these neoplasms is not problematic for FISH, which could detect abnormalities on cells in the interphase *(48,49)*.

5.1. CASE EXAMPLE OF APPLICATION OF FISH A 63-yr-old patient presented with a clinical history of vague abdominal discomfort and a 10-lb weight loss. A colonoscopy procedure showed the presence of numerous small colonic polyps, which were biopsied. Histological evaluation of the polyps revealed a dense, monotonous lymphoid infiltrate comprised of small to medium-sized lymphoid cells with ovoid to irregular, occasional cleaved nuclei, condensed chromatin, absent or indistinct nucleoli, and scant cytoplasm. By immunohistochemistry, the lymphoid infiltrate was diffusely positive for the B-cell antigen CD20 and coexpressed the T-cell antigen CD5. Kappa and lambda light-chain studies showed monoclonal lambda light-chain expression. An immunohistochemistry study to evaluate for cyclin D1 protein expression, characteristic of mantle cell lymphoma, showed only weak reactivity in rare cells and was thus inconclusive.

The morphologic and immunopathologic studies, including the presence of monoclonal kappa light-chain expression, were diagnostic of B-cell NHL. In particular, the findings were suspicious for mantle cell lymphoma; however, because of the equivocal study for cyclin D1, a definitive diagnosis of mantle cell lymphoma could not be established. The differential diagnosis also included B-cell chronic lymphocytic leukemia (CLL)/small lymphocytic lymphoma. Distinction between mantle cell lymphoma and B-cell CLL/small lymphocytic lymphoma is clinically important because mantle cell lymphoma is considered to be a clinically aggressive lymphoma and is associated with a poor prognosis (*see* Section 6.3).

To further evaluate for the possibility of mantle cell lymphoma, FISH was performed on paraffin sections *(50)*. Mantle cell lymphoma is associated with a specific chromosomal translocation, t(11;14)(q13;q32), which leads to overexpression of cyclin D1 protein. The result of the FISH procedure is shown in Fig. 15. The procedure utilizes a dual-color translocation probe system, which includes a probe directed to the immunoglobulin heavy-chain gene region (14q32) labeled with SpectrumGreen (green signal) and a probe directed to the *BCL-1/CCND1* gene region (11q13) labeled with SpectrumOrange (orange signal). A yellow

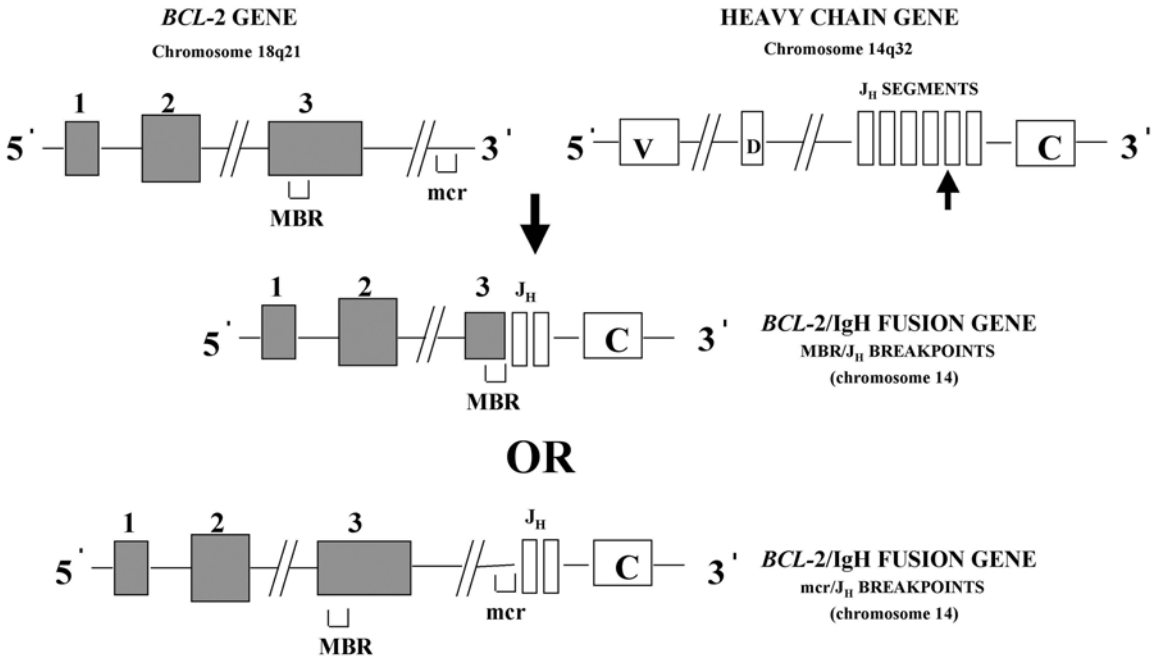


Fig. 16. Schematic diagram showing the translocation of *BCL-2* from chromosome 18 (shaded boxes) to the heavy-chain gene (IgH) on chromosome 14 (open boxes) resulting in a *BCL-2*/IgH fusion gene. For *BCL-2*, most breaks occur in either the MBR or mcr regions. Breakpoints in the heavy-chain gene involve J_H segments (arrowhead). The translocation may involve MBR and J_H breakpoints (middle panel) or mcr and J_H breakpoints (lower panel).

signal occurs because of fusion of the red and green signals and indicates the presence of a t(11;14)(q13;q32). The combined morphologic, immunopathologic, and molecular genetic findings in this biopsy are diagnostic of mantle cell lymphoma. The gastrointestinal presentation with multiple small polyps is indicative of a syndrome of multiple lymphomatous polyposis, which has been associated with mantle cell lymphoma.

6. CHROMOSOMAL TRANSLOCATIONS IN NON-HODGKIN'S LYMPHOMA

A number of nonrandom, reciprocal, and balanced translocations have been described in association with different subtypes of NHL *(9,51,52)*. These translocations often involve the transposition of a *proto-oncogene* from one chromosome to another. A proto-oncogene is defined as a normal cellular gene, that might be involved in the regulation of a variety of cellular processes, including cell growth, proliferation, differentiation, and apoptosis *(52)*. Characteristically, a proto-oncogene has the potential to contribute to neoplastic transformation when it is structurally or functionally altered, as occurs with a chromosomal translocation, a point mutation, or in the setting of gene amplification *(51)*. In NHL, reciprocal translocations frequently involve a proto-oncogene and the immunoglobulin or T-cell-receptor gene loci *(9,51,52)*.

Alterations in *tumor suppressor genes* are also involved in the pathogenesis of NHL. Tumor suppressor genes are defined as normal cellular genes that code for proteins that normally inhibit or suppress cell proliferation *(9,52)*. Inactivation of a tumor suppressor gene through the mechanisms of point mutation or gene deletion might contribute to uninhibited cell growth or proliferation. One example of a tumor suppressor gene is *p53*, which is the most frequently mutated gene in human cancers. In NHL, *p53* mutations are relatively common in high-grade, clinically aggressive lymphomas and relatively uncommon in low-grade, clinically indolent lymphomas *(9,52)*.

Chromosomal translocations can be demonstrated by molecular genetic methods, including Southern blotting, PCR, and FISH, as well as by traditional cytogenetic methods. Because cytogenetic methods are technically more difficult, especially when evaluating low-grade lymphoid neoplasms that are associated with a low mitotic rate, molecular genetic approaches are the method of choice. The demonstration of a specific chromosomal translocation might help characterize a specific type of lymphoma, especially when morphologic and immunopathologic studies are inconclusive (see Section 5.1). In addition, the identification of a specific translocation in a given case provides a marker that can be used to evaluate a patient for response to therapy and to monitor the patient for evidence of minimal residual disease. The following sections summarize several of the more common proto-oncogenes and their role in lymphomagenesis in several subtypes of NHL.

6.1. FOLLICULAR LYMPHOMA AND BCL-2 Follicular lymphoma is a relatively common B-cell NHL, that arises because of neoplastic transformation of cells derived from the lymphoid follicle or germinal center. Morphologically, follicular lymphoma is comprised of cells that resemble the normal cells of the germinal center, including small cleaved lymphocytes (or centrocytes) and large lymphoid cells (or centroblasts). The clinical behavior of follicular lymphoma is related to the number of small cleaved and large lymphoid cells comprising the neoplastic infiltrate. Follicular lymphoma with a

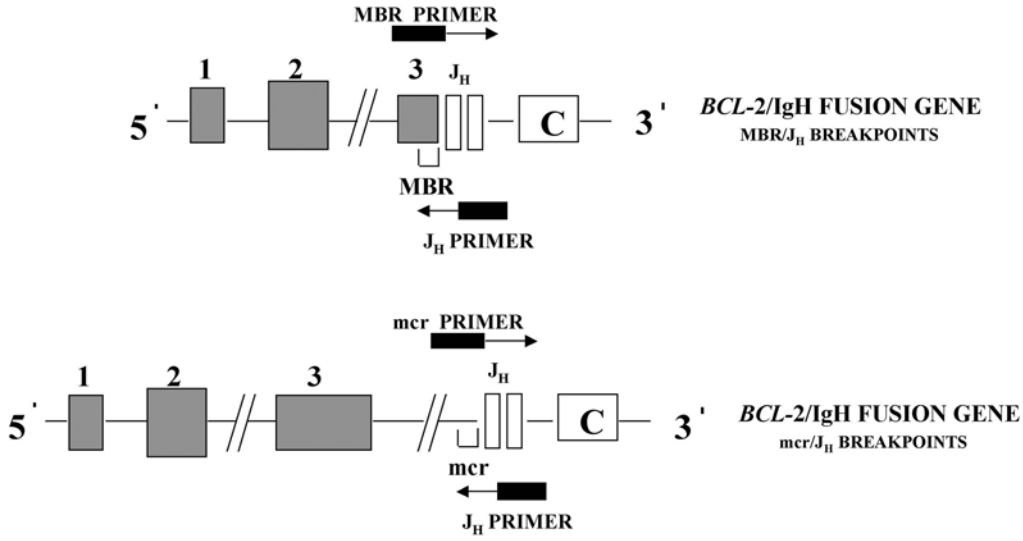


Fig. 17. Analysis for evidence of a *BCL-2*/IgH translocation by PCR involves two separate primer combinations: an MBR primer and consensus J_H primer (top), and an mcr and a consensus J_H primer (bottom).

predominance of small cleaved lymphocytes tends to be a relatively low-grade or indolent neoplasm, whereas follicular lymphoma with a higher proportion of large lymphoid cells tends to be clinically aggressive *(6)*. Follicular lymphoma is characteristically associated with a t(14;18)(q32q21), which occurs in approx 80–90% of cases *(53–55)*. This translocation could also be observed in up to 20% of cases of diffuse large B-cell lymphoma *(52)*. Similar to normal follicle center cells, the neoplastic cells in follicular lymphoma are associated with a high frequency of somatic hypermutation of the surface immunoglobulin receptor genes.

The t(14;18)(q32q21) results in the translocation of the *BCL-2* proto-oncogene, which normally resides on chromosome 18q21, to the B-cell heavy-chain gene locus on chromosome 14q32 *(53–55)*. Following translocation, the *BCL-2* proto-oncogene comes under direct control of immunoglobulin regulatory gene sequences. *BCL-2* is a part of a unique class of proto-oncogenes, which instead of promoting cell proliferation, functions by preventing apoptosis or programmed cell death *(53)*. In normal lymphocytes, expression of BCL-2 protein is limited to long-lived cells, including some subsets of normal T-cells and B-cells. Normal follicle center B-cells do not express BCL-2 protein. In follicular lymphoma, the t(14;18)(q32q21) results in upregulation of the *BCL-2* proto-oncogene and overexpression of BCL-2 protein. Overexpression of BCL-2 protein is likely one step in the process of lymphomagenesis, with elevated levels of BCL-2 extending the life span of neoplastic cells *(53–55)*.

Figure 16 schematically shows the reciprocal translocation involving the *BCL-2* locus on chromosome 18q21 and the immunoglobulin heavy-chain locus (IgH) on chromosome 14q32. The *BCL-2* gene contains three exons, including exon 1, which is a noncoding exon (upper panel, Fig. 16). The majority of chromosomal breaks occur in two regions referred to as the major breakpoint cluster region (MBR), where 50–75% of the breaks occur, and the minor breakpoint cluster region (mcr), where 20–40% of the breaks occur *(53–55)*. The MBR is located in exon 3 and the mcr is located downstream in the 3′ direction from exon 3. The breakpoints in the heavy-chain locus involve the J_H segments. The t(14;18) results in a *BCL-2*/IgH fusion gene as depicted in the middle and lower panels of Fig. 16.

Analysis for evidence of a *BCL-2*/IgH translocation can be performed by Southern blotting, PCR, and FISH. PCR is especially suited for analysis because the *BCL-2* and J_H breakpoints are located within a short segment of DNA *(55)*. Analysis by PCR is performed using two separate primer combinations to analyze for breaks at both the major and minor breakpoint cluster regions: a combination of MBR and J_H primers and a combination of mcr and J_H primers (Fig. 17). If a t(14;18) and, hence, a *BCL-2*/IgH rearrangement has not occurred, no PCR product will be obtained following amplification. Only when the primers are in close proximity to one another will a PCR amplified product be detected.

6.2. BURKITT LYMPHOMA AND C-MYC Burkitt lymphoma is a clinically aggressive, B-cell NHL characterized at the molecular level by a reciprocal translocation involving the *C-MYC* proto-oncogene, which normally resides on chromosome 8q24 *(56–67)*. Clinically, Burkitt lymphoma can occur in three different settings, including an *endemic* form, which occurs primarily in Africa and is associated with Epstein–Barr virus (EBV) infection in nearly 100% of cases, a *sporadic* form, which is seen throughout the world, and an *immunodeficiency-related* form, occurring primarily in human immunodeficiency virus (HIV) patients. A subset of both sporadic and immunodeficiency-related forms might also be associated with EBV infection *(6)*.

The most common translocation in Burkitt lymphoma is a t(8;14)(q24;q32), which results in the translocation of *C-MYC* to the B-cell heavy-chain gene locus on chromosome 14q32. This translocation occurs in approx 80% of Burkitt lymphoma cases regardless of the clinical setting. Other variant translocations

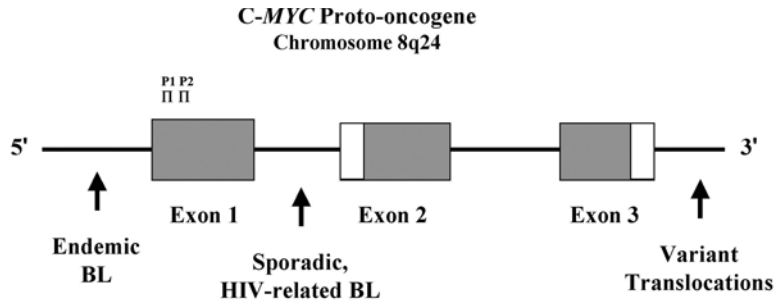


Fig. 18. Schematic diagram of *C-MYC* proto-oncogene on chromosome 8q24, which is comprised of three exons. Promoter regions (P) are in exon 1. Coding sequences are in exons 2 and 3 (shaded areas). Arrows indicate the breakpoints associated with various subtypes of Burkitt lymphoma.

involve the translocation of *C-MYC* to the kappa light-chain locus on chromosome 2, t(2;8)(p11;q24), which occurs in approx 15% of cases, and translocation of *C-MYC* to the lambda light-chain locus on chromosome 22, t(8;22)(q24;q11), which occurs in approx 5% of cases *(56–67)*.

The *C-MYC* gene encodes for C-MYC transcription factor, which is a basic helix–loop–helix zipper protein involved in the regulation of transcription *(62,67)*. The C-MYC transcription factor is normally involved in the transcriptional regulation of several target genes that play an important role in many aspects of cellular homeostasis, including cell cycle progression, cell proliferation, and apoptosis *(67)*. The t(8;14), t(2;8), or t(8;22) translocations in Burkitt lymphoma result in the juxtaposition of *C-MYC* to the immunoglobulin regulatory sequences with consequent upregulation of *C-MYC* transcription and overexpression of C-MYC transcription factor. The overexpression of C-MYC transcription factor results in enhanced cell cycle progression and cell proliferation. In Burkitt lymphoma, cellular proliferation is among the highest of any human tumor.

The *C-MYC* proto-oncogene is comprised of three exons (Fig. 18). Exon 1 contains the two major promoter regions of *C-MYC*, which are DNA sequences where RNA polymerase binds and initiates transcription. Exons 2 and 3 contain the coding sequences for C-MYC transcription factor. In Burkitt lymphoma, the t(8;14) breakpoints are quite varied in both chromosome 8 and in chromosome 14. In endemic Burkitt lymphoma, the chromosome 8 breakpoint is located a significant distance 5′ to the *C-MYC* gene locus. In contrast, in sporadic and HIV-related Burkitt lymphoma, the chromosome 8 breakpoints tend to occur between exons 1 and 2 of the *C-MYC* gene *(63)*. With respect to variant translocations involving the immunoglobulin kappa and lambda light-chain gene loci and *C-MYC*, the breakpoints on chromosome 8 are located a variable distance 3′ of the locus *(67)*. Because of the variability of breakpoints involving the immunoglobulin and C-*MYC* genes, PCR approaches for detecting the various translocations are not practical. Detection of the various translocations is best accomplished in the clinical laboratory setting by either FISH or traditional karyotyping of metaphase chromosomes.

6.3. MANTLE CELL LYMPHOMA AND BCL-1 Mantle cell lymphoma is a pathologically distinct B-cell NHL characterized at the molecular level by a translocation, t(11;14)(q13;q32), involving the *BCL-1* (B-cell leukemia/lymphoma 1) gene region on chromosome 11q13 and the immunoglobulin heavy-chain gene locus on chromosome 14q32 *(68–81)*. Histologically, mantle cell lymphoma is characterized by a monomorphous proliferation of cytologically atypical, small to medium-sized lymphoid cells *(6)*. In contrast to other "small" B-cell lymphomas, such as follicular lymphoma and marginal zone B-cell lymphoma, which are usually associated with an indolent clinical course and prolonged survival, mantle cell lymphoma is a clinically aggressive lymphoma *(69)*. Most cases of mantle cell lymphoma are associated with a poor response to conventional chemotherapeutic regimens and a relatively short median survival of only 3–4 yr *(6)*. The "blastoid" variant of mantle cell lymphoma is characterized by a proliferation of medium to large lymphoid cells and an even more aggressive clinical course *(6)*.

The 11q13 gene region includes both the *BCL-1* and the *CCND1* (cyclin D1) gene loci (Fig. 19). The *CCND1* gene locus was originally designated PRAD1 because of its initial detection in parathyroid adenomas and is positioned downstream from the *BCL-1* locus *(81)*. The breaks in the 11q13 gene region are widely scattered within the *BCL-1* locus; however, approx 30–40% of breaks occur within the MTC (major translocation cluster) region of the *BCL-1* locus *(76–78)*. The CCND1 coding region is unaffected by breaks in the *BCL-1* locus; thus remaining structurally intact. Following translocation, the *CCND1* gene comes under direct control of the immunoglobulin regulatory sequences and is upregulated. The *CCND1* gene encodes a cyclin D1 protein, which is not expressed in normal lymphoid cells but is overexpressed following the t(11;14) (79-81).

Cyclin D1 is one of several proteins involved in cell cycle regulation, which, in turn, is essential for the control of cellular proliferation and differentiation. The G_1 to S checkpoint of the cell cycle, in particular, is under the balanced control of several regulatory proteins, which either promote or inhibit progression of the cell cycle. Cyclin D1 is a positive cell-cycle-regulating protein that interacts with important enzymes termed *cyclin-dependent kinases* (CDKs), which promote progression of cells from the G_1-phase to the S-phase. The overexpression of cyclin D1 following the translocation t(11;14)(q13;q32) drives cells to proliferate and is important in the pathogenesis of mantle cell lymphoma *(79–81)*. A minority of cases of mantle cell lymphoma, including the blastoid variant, are also associated with abnormalities of negative cell cycle regulatory proteins, such as p16/18 and p53 (tumor suppressor genes), which normally

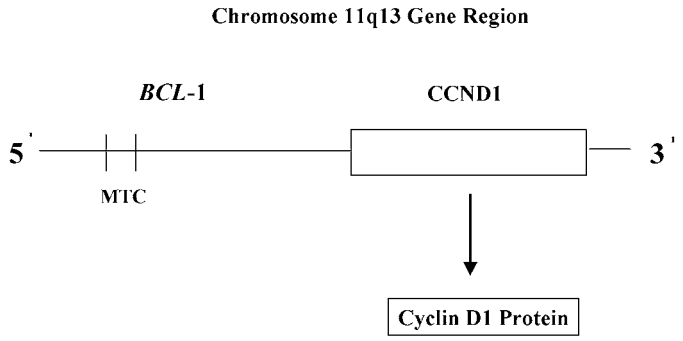


Fig. 19. Schematic diagram of the chromosome 11q13 gene region, which includes the *BCL-1* locus and the *CCND1* gene. The breaks in the 11q13 gene region are widely scattered within the *BCL*-1 locus; however, approx 30–40% of breaks occur within the MTC (major translocation cluster) region of the *BCL-1* locus. The *CCND1* gene encodes for the cyclin D1 protein.

inhibit cell cycle progression and further drive cells to proliferate *(70,79–81)*.

In the clinical laboratory, the demonstration of either cyclin D1 overexpression by immunohistochemistry or a translocation t(11;14) by molecular genetic methods is essential to establishing a diagnosis of mantle cell lymphoma. Molecular genetic approaches are used primarily on fresh tissue samples such as peripheral blood and bone marrow aspirate specimens, but they could also be used on paraffin-embedded tissue sections. PCR approaches utilizing primers directed to the immunoglobulin heavy-chain J_H segments and the major cluster region (MCR) of the *BCL-1* gene are useful, however, this method will detect the *bcl-1* translocation in only 30–40% of cases *(70)*. The *BCL-1* translocation can also be detected by Southern blot using multiple probes, however, this is a labor-intensive, costly process. More recently, FISH has been shown to be highly sensitive method for demonstrating the *BCL-1* translocation in both paraffin-embedded tissue and fresh tissue *(50)*.

6.4. DIFFUSE LARGE B-CELL LYMPHOMA AND BCL-6 Diffuse large B-cell lymphoma is one of the more common lymphomas in Western countries, comprising approx 30–40% of all lymphomas *(6)*. Clinically, diffuse large B-cell lymphoma occurs primarily in adults, but could occur in any age group and is considered to be an aggressive, but potentially curable lymphoma *(6,69)*. A variety of different chromosomal alterations could be observed in diffuse large B-cell lymphoma, which could lead to the presence of complex karyotypes.

One translocation observed in up to 30–40% of diffuse large B-cell lymphomas involves the *BCL-6* proto-oncogene, which normally resides on chromosome 3q27 *(82–89)*. Chromosome alterations involving the 3q27 gene region could involve at least 10 different partner chromosomes, including the heavy-chain gene locus (14q32), the kappa light-chain gene locus (2p12), and the lambda light-chain gene locus (22q11). *BCL-6* encodes a zinc-finger transcription factor (BCL-6), which is involved in the process of germinal center B-cell differentiation *(82,83)*. In normal tissues, BCL-6 protein expression is restricted to germinal center B-cells and is reduced or absent in other stages of B-cell development. The various translocations involving 3q27 result in deregulation of *BCL-6* gene expression, which likely contributes to lymphomagenesis.

The *BCL-6* gene is comprised of 10 exons and the majority of breaks are tightly clustered within the first intron or exon *(87–89)*, referred to as the major translocation cluster region. *BCL-6* translocations can thus be detected with greatest efficiency by Southern blot analysis using a DNA probe directed to DNA sequences just downstream of the major translocation cluster region.

6.5. ANAPLASTIC LARGE-CELL LYMPHOMA AND NPM/ALK Anaplastic large-cell lymphoma (ALCL) is a clinically, morphologically, and genetically heterogenous T-cell NHL, which is characterized in many cases at the molecular level by a translocation involving the anaplastic lymphoma kinase (*ALK*) proto-oncogene *(90–95)*. A number of morphologic variants of ALCL have been described, including the common variant, the small-cell variant, and the lymphohistiocytic variant. *(6,93,95)*. All morphologic variants of ALCL include the so-called "hallmark cell," which is a large, pleomorphic (anaplastic) cell, frequently with a lobulated, horseshoe, or kidney-shaped nucleus *(6,93,95)*. The neoplastic cells are characteristically positive for the CD30 antigen. ALCL can be broadly divided into ALK-positive and ALK-negative subtypes based on the presence or absence of a translocation involving the *ALK* proto-oncogene and resultant expression of the ALK protein product (Fig. 1B). ALK-positive ALCL tends to occur in younger individuals, usually in the first three decades of life, whereas ALK-negative ALCL tends to occur in older individuals. Despite the histologically aggressive appearance of ALCL, the presence of ALK positivity in ALCL confers one of the most clinically favorable prognostic groups with a very high overall cure rate. In contrast, ALK-negative ALCL is associated with a less favorable prognosis *(6,93,95)*.

The most common translocation in ALCL is a t(2;5)(p23;q35), which occurs in approx 75% of cases of ALK-positive ALCL. This translocation results in a reciprocal translocation involving the *ALK* gene, which normally resides on chromosome 2p23 and the nucleophosmin (*NPM*) gene, which normally resides on chromosome 5q35 *(91,92,94,95)*. Other less common or variant translocations in ALK-positive ALCL involve the *ALK* gene on chromosome 2p23 and chromosomes 1, 2, 3, and 22 *(94,95)*.

The wild-type ALK protein is a transmembrane receptor, which includes a tyrosine kinase domain. Normal expression of ALK is restricted to some neural cells; normal lymphocytes do not express ALK. Wild-type nucleophosmin is a ubiquitous protein, which is normally located in the nucleolus and plays a role in ribosome assembly by shuttling ribonuclear complexes between the nucleus and cytoplasm *(94,95)*. The t(2;5)(p23;q35) juxtaposes the *NPM* and *ALK* genes on the derived chromosome 5 and results in a *NPM/ALK* chimeric gene, which codes for an 80-kDa, NPM/ALK fusion protein, designated p80 (Fig. 20). The upregulated NPM/ALK fusion protein includes the preserved ALK tyrosine kinase domain, which likely contributes to the dysregulation of several downstream signals, thus promoting lymphomagenesis *(94,95)*.

In patients with ALCL, the demonstration of a t(2;5) by molecular genetic methods or ALK overexpression by immunopathological methods is important for both diagnostic and prognostic

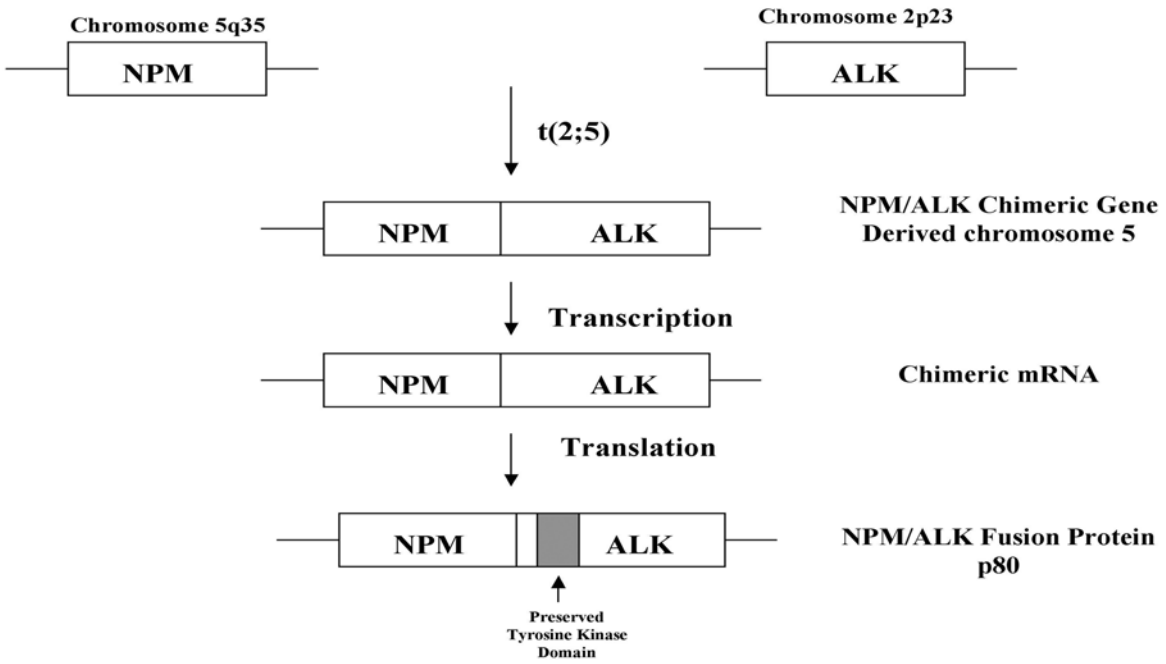


Fig. 20. Schematic diagram of the t(2;5)(p23;q35). This translocation juxtaposes the *NPM* and *ALK* genes on the derived chromosome 5, resulting in a *NPM/ALK* chimeric gene, which codes for an 80-kda, NPM/ALK fusion protein (designated p80). The NPM/ALK fusion protein includes a preserved tyrosine kinase domain (shaded area).

purposes. Immunohistochemistry is the preferred method because this method will detect ALK protein in cases of ALCL with the t(2;5), as well as in cases of ALCL with variant translocations. In contrast, most molecular genetic methods are primarily directed to the t(2;5). Reverse transcriptase (RT)–PCR is one molecular genetic approach routinely used in the clinical laboratory. Because the breakpoints on chromosomes 2 and 5 are widely separated, direct PCR methods are generally unsuccessful. To circumvent this, the RT-PCR approach targets *NPM/ALK* mRNA. Using reverse transcriptase, a complementary DNA strand is produced from the *NPM/ALK* mRNA. The resultant cDNA has primer annealing sequences in closer proximity, thus allowing for direct PCR analysis *(94,95)*.

7. HODGKIN'S LYMPHOMA

Hodgkin's lymphoma is the second major category of lymphoma, which accounts for approx 14% of all lymphomas annually in the United States. Hodgkin's lymphoma presents most commonly in lymph nodes, especially in the cervical region; extranodal presentations of Hodgkin's lymphoma are relatively uncommon. Although Hodgkin's lymphoma can occur at any age, most cases occur in young adults *(3,6)*.

The diagnosis of Hodgkin's lymphoma is established based on the morphologic identification of neoplastic Reed–Sternberg cells (RS cells) and Reed–Sternberg variants (RS variants) in the appropriate benign inflammatory cell background *(96–101)*. In most cases of Hodgkin's lymphoma, RS cells and RS variants comprise less than 1–2% of the total cell population; the bulk of the infiltrate is comprised of benign inflammatory cells. Immunopathologic studies are also routinely used to confirm the diagnosis of Hodgkin's lymphoma and to distinguish Hodgkin's lymphoma from non-Hodgkin's lymphoma *(97,98)*.

Based on the WHO Classification, Hodgkin's lymphoma can be divided into two groups: classical Hodgkin's lymphoma and nodular lymphocyte predominant Hodgkin's lymphoma *(6)*. The distinction between the two groups is based on the morphologic and immunopathologic characteristics of the RS cells and RS variants. In classical Hodgkin's lymphoma, the RS cells and RS variants are typically positive for the myeloid antigen CD15 and the activation antigen CD30, but are negative for B-cell antigens, including CD20. In contrast, in nodular lymphocyte predominant Hodgkin's lymphoma, the RS variants (referred to as L&H cells) are negative for CD15 and CD30, but strongly positive for CD20 *(6,97,98)*.

The cell of origin of the RS cell had been an enigma for decades; however, recent immunopatholgic and sophisticated molecular genetic studies have clarified the nature of the RS cell *(96–101)*. In particular, the development and refinement of methods that allow for isolation of individual RS cells by micromanipulation with subsequent molecular genetic analysis have been most informative *(97,98)*. The analysis of individual RS cells using primers directed to immunoglobulin V_H segments has confirmed that the RS cells in virtually all cases of nodular lymphocyte predominant Hodgkin's lymphoma and classical Hodgkin's lymphoma are of B-cell origin and are monoclonal. In addition, these studies have demonstrated a high rate of somatic hypermutation in RS cells, suggesting an origin from germinal center B-cells *(101)*.

Despite the B-cell origin of RS cells, Southern blot methods for determining clonality on whole tissue sections contribute very little to the routine diagnosis of Hodgkin's lymphoma because the RS cells and variants comprise only 1–2% of all cells in the tissue sample and the sensitivity of the Southern blot is approx 5% *(100)*. Routine PCR methods performed on

whole tissue sections are more sensitive than the Southern blot, but they are also of very limited use because monoclonal rearrangements are detected in only a small number of cases.

8. DETECTION OF MINIMAL RESIDUAL DISEASE

An increasingly important application of molecular genetic testing in both lymphoma and leukemia patients involves the monitoring of patient response to various therapies. A patient is characterized as being in *complete clinical remission* following therapy when there is no detectable disease based on standard clinical, radiological, and laboratory evaluations *(102)*. However, a sizable fraction of patients considered to be in complete clinical remission will eventually relapse and die of their disease. Disease relapse in these patients is presumably caused by the presence of *minimal residual disease* (MRD), which refers to the presence of persistent small numbers of lymphoma or leukemia cells that are below the level of detection of standard clinical and laboratory evaluations *(102–107)*.

A variety of different techniques have been devised to assess patients for evidence of MRD including traditional morphology, immunohistochemistry, flow cytometry, cell culture methods, and conventional cytogenetic methods, as well as molecular genetic methods, including FISH, Southern blotting, and PCR *(107)*. Each of these methods has advantages and disadvantages. With the exception of flow cytometry and PCR, most of the approaches listed lack sensitivity and are capable of detecting only a 1–5% malignant cell population (i.e., 1–5 malignant cells/100 normal cells). In contrast, multiparameter flow cytometry has the capability of detecting a few malignant cells among 10^3 or 10^4 normal cells, and PCR approaches have the capability of detecting a few malignant cells among 10^5 or 10^6 normal cells *(102,107)*.

Evaluation for evidence of MRD in NHL has focused on B-cell NHL, which tend to present in a disseminated fashion with frequent involvement of the peripheral blood and bone marrow *(102–107)*. These lymphomas include B-cell CLL/small lymphocytic lymphoma, follicular lymphoma, and mantle cell lymphoma. In order to evaluate for the presence of MRD by PCR, the malignant cells must have a unique set of DNA sequences, which allow for distinction from normal cells. PCR evaluation for evidence of a specific chromosomal translocation, such as the t(14;18)(q32;q21) in patients with follicular lymphoma and the t(11;14)(q13;q32) in patients with mantle cell lymphoma, provides an ideal marker of MRD. Patients can also be monitored for MRD by evaluating for evidence of a B-cell heavy-chain (IgH) rearrangement using clone-specific primers directed to V_H and J_H segments. Patients without detectable MRD by molecular methods are considered to be in *molecular remission (102).*

The ability to detect and quantitate MRD by molecular methods would be expected to improve clinical management by optimizing therapies. In some cases, the detection of MRD might indicate a high risk for clinical relapse and warrant the use of specific therapies, such as modulation of graft-vs-host disease, that are most efficacious in patients with small residual tumor burdens *(102)*. Alternatively, some patients might require more aggressive therapeutic approaches such as bone marrow transplantation. In contrast, patients in both clinical and molecular remission might require less intensive therapy, thus reducing patient exposure to potentially toxic drugs.

9. CONCLUSION

Our understanding of the malignant lymphomas has evolved tremendously over the last two decades. The introduction of immunopathologic techniques in the 1980s both refined and redefined both non-Hodgkin's lymphoma and Hodgkin's lymphoma. The NHLs, in particular, were closely related to the normal immune system. The application of cytogenetic and molecular genetic technology since the mid 1980s has further identified specific, nonrandom chromosomal alterations in many lymphomas, thus further defining many of these neoplasms as specific pathologic entities.

The routine application of molecular genetic techniques has significantly enhanced our ability to precisely diagnose and classify lymphoid neoplasms. A definitive diagnosis of NHL might only be established after the molecular genetic demonstration of a clonal B- or T-cell gene rearrangement. In addition, the demonstration of a specific chromosomal translocation might help define a specific subtype of lymphoma. A clonal B- or T-cell gene rearrangement or a specific translocation might also serve as molecular marker to evaluate patients for evidence of MRD.

Molecular genetic techniques have also contributed to our understanding of the pathogenesis of many lymphomas. The chromosomal translocations often involve the transposition of a specific proto-oncogene, which might alter normal cellular processes such as cell growth, proliferation, and apoptosis. Further understanding of the molecular mechanisms involved in lymphomagenesis will likely lead to more specific therapies targeted at molecular processes.

REFERENCES

1. Hodgkin, T. On some morbid appearances of the absorbent glands and spleen. *Med–Chir. Trans.* **17:**69–97, 1832.
2. Wilks, S. Cases of enlargement of the lymphatic glands and spleen (or, Hodgkin's disease), with remarks. *Guys Hosp. Rep.* **11:**56–67, 1865.
3. Jemal, A., Murray, T., Samuels, A., et al. Cancer statistics, 2003. *CA Cancer J. Clin* **53:**5–26, 2003.
4. Bertoli, L. F., and Burrows, P. D. Normal B-lineage cells: their differentiation and identification. *Clin. Lab. Med.* **8:**15–30, 1988.
5. Denning, S. M. and Haynes, B. F. Differentiation of human T cells. *Clin. Lab. Med.* **8:**1–14, 1988.
6. Jaffe, E. S., Harris, N. L., Stein, H., and Vardiman, J. eds. *World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of Haematopoeitic and Lymphoid Tissues.* IARC, Lyon, 2001.
7. Jaffe, E. S. The role of immunophenotypic markers in the classification of non-Hodgkin's lymphomas. *Semin. Oncol.* **17:**11–19, 1990.
8. Campana, D. and Pui, C-H. Detection of minimal residual disease in acute leukemia: methodologic advances and clinical significance. *Blood* **85:**1416–1434, 1995.
9. Sen, F., Vega, F, and Medeiros, L. J. Molecular genetic methods in the diagnosis of hematologic neoplasms. *Semin. Diagn. Pathol.* **19(2):**, 72–93, 2002.
10. Cooper, M. D. B lymphocytes: normal development and function. *N. Engl. J. Med.* **317:**1452–1456, 1987.
11. Pan, L., Cesarman, E., and Knowles, D. M. Antigen receptor genes: structure, function, and genetic analysis of their rearrangements, *Neoplastic Hematopathology,* Knowles, D. M., ed., Lippincott, Williams and Wilkins, Philadelphia, PA, pp. 307–328, 2001.
12. Foon, K. A. and Todd, R. F. Immunologic classification of leukemia and lymphoma. *Blood* **68:**1–31, 1986.

13. Spits, H., Lanier, L. L., and Phillips, J. H. Development of human T and natural killer cells. *Blood* **85:**2654–2670, 1995.
14. Cossman, J., Uppenkamp, M., Sundeen, J., Coupland, R., and Raffeld, M. Molecular genetics and the diagnosis of lymphoma. *Arch. Pathol. Lab. Med.* **112:**117–127, 1988.
15. Cossman, J., and Uppenkamp, M. T cell gene rearrangements and the diagnosis of T cell neoplasms. *Clin. Lab. Med.* **8:**31–44, 1988.
16. Royer, H. D. and Reinherz, E. L. T lymphocytes: ontogeny, function, and relevance to clinical disorders. *N. Eng. J. Med.* **317:** 1136–1142, 1987.
17. Gill, J. I. and Gulley, M. L. Immunoglobulin and T cell receptor gene rearrangement. *Hematol. Oncol. Clin. North Am.* **8:**751–770, 1994.
18. Rezuke, W. N., Abernathy, E. C., and Tsongalis, G. J. Molecular diagnosis of B and T cell lymphomas: fundamental principles and clinical applications. *Clin. Chem.* **43:**1814–1823, 1997.
19. Cossman, J., Zehnbauer, B., Garrett, C. T., et al. Gene rearrangements in the diagnosis of lymphoma/leukemia. Guidelines for use based on a multiinstitutional study. *Am. J. Clin. Pathol.* **95:**347–354, 1991.
20. Farkas, D. H. The Southern blot: application to the B and T cell gene rearrangement test. *Lab. Med.* **23:**723–729, 1992.
21. Beishuizen, A., Verhoeven, M. J., Mol, E. J., and van Dongen, J. J. M. Detection of immunoglobulin kappa light chain gene rearrangement patterns by Southern blot analysis. *Leukemia* **8:**2228–2236, 1994.
22. Hodges, K. A., Kosciol, C. M., Rezuke, W. N., Abernathy, E. C., Pastuszak, W. T., and Tsongalis, G. T. Chemiluminescent detection of gene rearrangements in hematologic malignancy. *Ann. Clin. Lab. Sci.* **26:**114–118, 1996.
23. Medeiros, L. J. and Weiss, L. M. The utility of the polymerase chain reaction as a screening method for the detection of antigen receptor gene rearrangements. *Hum. Pathol.* **25:**1261–1263, 1994.
24. Macintyre, E. A. The use of the polymerase chain reaction in hematology. *Blood Rev.* **3:**201–210, 1989.
25. Weiss, L. M. and Spagnolo, D. V. Assessment of clonality in lymphoid proliferations. *Am. J. Pathol.* **142:**1679–1682, 1993.
26. McCarthy, K. P., Sloane, J. P., Kabarowski, J. H. S., Matutes, E., and Wiedemann, L. M. The rapid detection of clonal T-cell proliferations in patients with lymphoid disorders. *Am. J. Pathol.* **138:**821–828, 1991.
27. Ben-Ezra, J. Variable rate of detection of immunoglobulin heavy chain V-D-J rearrangement by PCR: a systematic study of 41 B-cell non-Hodgkin's lymphomas and leukemias. *Leuk. Lymph.* **7:**289–295, 1992.
28. Diss, T. C., Peng, H., Wotherspoon, A. C., Isaacson, P. G., and Pan, L. Detection of monoclonality in low-grade B-cell lymphomas using the polymerase chain reaction is dependent on primer selction and lymphoma type. *J. Pathol.* **169:**291–295, 1993.
29. Segal, G. H., Jorgensen, T., Masih, A. S., and Braylan, R. C. Optimal primer selection for clonality assessment by polymerase chain reaction analysis. Low grade B-cell lymphoproliferative disorders of nonfollicular center cell type. *Hum. Pathol.* **25:**1269–1275, 1994.
30. Slack, I. N., McCarthy, K. P., Wiedemann, L. M., and Sloane, J. P. Evaluation of sensitivity, specificity, and reproducibility of an optimized method for detecting clonal rearrangements of immunoglobulin and T-cell receptor genes in formalin-fixed, paraffin-embedded sections. *Diagn. Mol. Pathol.* **2:**223–232, 1993.
31. Algara, P., Caridad, S., Martinez, P., et al. Value of PCR detection of TCRγ gene rearrangments in the diagnosis of cutaneous lymphocytic infiltrates. *Diagn. Mol. Pathol.* **3:**275–282, 1994.
32. Achille, A., Scarpa, A., Montresor, M., et al. Routine application of polymerase chain reaction in the diagnosis of monoclonality of B-cell lymphoid proliferations. *Diagn. Mol. Pathol.* **4:**14–24, 1995.
33. Fodinger, M., Buchmayer, H., Schwarzinger, I., et al. Multiplex PCR for rapid detection of T-cell receptor gamma chain gene rearrangements in patients with lymphoproliferative diseases. *Br. J. Haematol.* **94:**136–139, 1996.
34. Lozano, M. D., Tierens, A., Greiner, T. C., et al. Clonality analysis of B-lymphoid proliferations using the polymerase chain reaction. *Cancer* **77:**1349–1355, 1996.
35. Abdel-Reheim, F. A., Edwards, E., and Arber, D. A. Utility of a rapid polymerase chain reaction panel for the detection of molecular changes in B-cell lymphoma. *Arch. Pathol. Lab. Med.* **120:**357–363, 1996.
36. Lombardo, J. F., Hwang, T. S., Maiese, R. L., Millson, A., and Segal, G. H. Optimal primer selection for clonality assessment by polymerase chain reaction analysis: Intermediate and high grade B-cell neoplasms. *Hum. Pathol.* **27:**373–380, 1996.
37. Diaz-Cano, S. PCR-based alternative for diagnosis of immunoglobulin heavy chain gene rearrangement. *Diagn. Mol. Pathol.* **5:**3–9, 1996.
38. Ritter, J. H., Wick, M. R., Adesokan, P. N., et al. Assessment of clonality in cutaneous lymphoid infiltrates by polymerase chain reaction analysis of immunoglobulin heavy chain gene rearrangements. *Am. J. Clin. Pathol.* **108:**60–68, 1997.
39. Ashton-Key, M., Diss, T. C., Du, M. Q., et al. The value of the polymerase chain reaction in the diagnosis of cutaneous T-cell infiltrates. *Am. J. Surg. Pathol.* **21:**743–747, 1997.
40. Tsongalis, G. J., Stevenson, A. J., Hodges, K. A., and Rezuke, W. N. Clonal analysis of B-cell lymphoproliferative disorders. *Eur. J. Lab. Med.* **3:**142–145, 1998.
41. Sioutos, N., Bagg, A., Michaud, G. Y., et al. Polymerase chain reaction versus Southern blot hybridization. Detection of immunoglobulin heavy-chain gene rearrangements. *Diagn. Mol. Pathol.* **4:**8–13, 1995.
42. Lehman, C. M., Sarago, C., Nasim, S., Comerford, J., Karcher, D. S., and Garrett, C. T. Comparison of PCR with Southern hybridization for the routine detection of immunoglobulin heavy chain gene rearrangements. *Am. J. Clin. Pathol.* **103:**171–176, 1995.
43. Tbakhi, A. and Tubbs, R. R. Utility of the polymerase chain reaction in detecting B cell clonality in lymphoid neoplasms. *Cancer* **77:**1223–1225, 1996.
44. Kuppers, R., Zhao, M., Rajewsky, K., and Hansmann, M. L. Detection of clonal B cell populations in paraffin-embedded tissues by polymerase chain reaction. *Am. J. Pathol.* **143:**230–239, 1993.
45. Inghirami, G., Szabolcs, M. J., Yee, H. T., Corradini, P., Cesarman, E., and Knowles, D. M. Detection of immunoglobulin gene rearrangement of B cell non-Hodgkin's lymphomas and leukemias in fresh, unfixed and formalin fixed, paraffin-embedded tissue by polymerase chain reaction. *Lab. Invest.* **68:**746–757, 1993.
46. Reed, T. J., Reid, A., Wallberg, K., O'Leary, T. J., and Frizzera, G. Determination of B cell clonality in paraffin-embedded lymph nodes using the polymerase chain reaction. *Diagn. Mol. Pathol.* **2:**42–49, 1993.
47. Segal, G. H., Jorgensen, T., Scott, M., and Braylan, R. C. Optimal primer selection for clonality assessement by polymerase chain reaction analysis: II. Follicular lymphomas. *Hum. Pathol.* **25:** 1276–1282, 1994.
48. Siebert, R. and Weber-Matthiesen, K. Fluorescence in situ hybridization as a diagnostic tool in malignant lymphomas. *Histochem. Cell. Biol.* **108:**391–402, 1997.
49. Kluin, P. H. and Schuuring, E. FISH and related techniques in the diagnosis of lymphoma. *Cancer Surv.* **30:**3–20, 1997.
50. Remstein, E. D., Kurtin, P. J., Buno, I., et al. Diagnostic utility of fluorescence in situ hybridization in mantle cell lymphoma. *Br. J. Haematol.* **110:**856–862, 2000.
51. Medeiros, L. J. and Carr, J. Overview of the role of molecular methods in the diagnosis of malignant lymphomas. *Arch. Pathol. Lab. Med.* **123:**1189–1207, 1999.
52. Tam, W. and Dalla-Favera, R. Protooncogenes and tumor suppressor genes in hematopoietic malignancies, in *Neoplastic Hematopathology,* Knowles, D. M., ed., Lippincott, Williams and Wilkins, Philadelphia, PA, 2001, pp. 329–364.
53. Korsmeyer, S. J. Bcl-2 initiates a new category of oncogenes: regulators of cell death. *Blood* **80:**879–886, 1992.
54. Crisan, D., Chen, S-T., and Weil, S. C. Polymerase chain reaction in the diagnosis of chromosomal breakpoints. *Hematol. Oncol. Clin. North. Am.* **8:**725–750, 1994.
55. Ngan, B-Y., Nourse, J., and Cleary, M. L. Detection of chromosomal translocation t(14;18) within the minor cluster region of *bcl-2*

by polymerase chain reaction and direct genomic sequencing of the enzymatically amplified DNA in follicular lymphomas. *Blood* **73:**1759–1762, 1989.
56. Cory, S., Activation of cellular oncogenes in hemopoietic cells by chromosome translocation. *Adv. Cancer Res.* **47:**189–234, 1986.
57. Adams, J. M., Gerondakis, S., Webb, E., Mitchell, J., Bernard, O., and Cory, S. Transcriptionally active DNA region that rearranges frequently in murine lymphoid tumors. *Proc. Natl. Acad. Sci. USA* **79:**6966–6970, 1982.
58. Adams, J. M., Gerondakis, S., Webb, E., Corcoran, L. M., and Cory, S. Cellular myc oncogene is altered by chromosome translocation to an immunoglobulin locus in murine plasmacytomas and is rearranged similarly in human Burkitt lymphomas. *Proc. Natl. Acad. Sci. USA* **80:**1982–1986, 1983.
59. Taub, R., Kirsch, I., Morton, C., et al. Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt and murine plasmacytoma cells. *Proc. Natl. Acad. Sci. USA* **79:**7837–7841, 1982.
60. Dalla-Favera, R., Martinotti, S., Gallo, R. C., Erikson, J., and Croce, C. M. Translocation and rearrangements of the c-myc oncogene locus in human undifferentiated B cell lymphomas. *Science* **219:**963–967, 1983.
61. Bernard, O. Cory, S., Gerondakis, S., Webb, E., and Adams, J. M. Sequence of the murine and human cellular myc oncogenes and two modes of myc transcription resulting from chromosome translocation in B lymphoid tumors. *EMBO J.* **2:**2375–2383, 1983.
62. Cory, S., Vaux, D. L., Strasser, A., Harris, A. W., and Adams, J. M. Insights from Bcl-2 and myc: malignancy involves abrogation of apoptosis as well as sustained proliferation. *Cancer Res.* **59:**1685s–1692s, 1999.
63. Yano, T., Sander, C. A., Clark, H. M., Dolezal, M. V., Jaffe, E. S., and Raffeld, M. Clustered mutations in the second exon of the myc gene in sporadic Burkitt's lymphoma. *Oncogene* **8:**2741–2748, 1993.
64. Cheng, S. W. G., Davies, K. P., Yung, E., Beltran, R. J., Yu, J., and Kalpana, G.V. C-myc interacts with INI1/hSNF5 and requires the SWI/SNF complex for transactivation function. *Nat. Genet.* **22:**102–105, 1999.
65. Facchini, L.M. and Penn, l. Z. The molecular role of myc in growth and transformation: recent discoveries lead to new insights. *FASEB J.* **12:**633–651, 1998.
66. Henriksson, M. and Luscher, B. Proteins of the myc network: essential regulators of cell growth and differentiation. *Adv. Cancer Res.* **68:**109–182, 1996.
67. Hecht, J. L. and Aster, J. C. Molecular biology of Burkitt's lymphoma. *J. Clin. Oncol.* **18:**3707–3721, 2000.
68. Williams, M. E., Westermann, C. D., and Swerdlow, S. H. Genotypic characterization of centrocytic lymphoma: frequent rearrangement of the chromosome 11 bcl-1 locus. *Blood* **76:**1387–1391, 1990.
69. Armitage, J. O. A clinical evaluation of the International Study Group Classification of non-Hodgkin's lymphoma. The Non-Hodgkin's Lymphoma Classification Project. *Blood* **89(11):** 3909–3918, 1997.
70. Fiel-Gan, M. D., Almeida, L., Rose, D. C., et al. Proliferative fraction, bcl-1 gene translocation, and p53 mutation status as markers in mantle cell lymphoma. *Int. J. Mol. Med.* **3:**373–379, 1999.
71. Seto, M., Yamamoto, K., Iida, S., et al. Gene rearrangement and overexpression of PRAD-1 in lymphoid malignancy with t(11;14)(q13;q32) translocation. *Oncogene* **7:**1401–1406, 1992.
72. Motokura, T., Bloom, T., Kim, H. G., et al. A novel cyclin encoded by bcl-1 linked candidate oncogene. *Nature* **350:**512–515, 1991.
73. Argatoff, L. H., Connors, J. M., Klasa, R. J., Horsman, D. E., and Gascoyne, R. D. Mantle cell lymphoma: a clinicopathologic study of 80 cases. *Blood* **89:**2067–2078, 1997.
74. Chibbar, R., Leung, K., McCormick, S., et al. bcl-1 gene rearrangements in mantle cell lymphoma: a comprehensive analysis of 118 cases, including B-5 fixed tissue, by polymerase chain reaction and Southern transfer analysis. *Mod. Pathol.* **11:**1089–1097, 1998.
75. Molot, R. J., Meeker, T. C., Wittwer, C. T., et al. Antigen expression and polymerase chain reaction amplification of mantle cell lymphomas. *Blood* **83:**1626–1631, 1994.
76. Williams, M. E., Swerdlow, S. H., Rosenberg, C. L., and Arnold, A. Characterization of chromosome 11 translocation breakpoints at the bcl-1 and PRAD1 loci in centrocytic lymphoma. *Cancer Res.* **52:**5541s–5544s, 1992.
77. Williams, M. E., Meeker, T. C., and Swerdlow, S. H. Rearrangement of the chromosome 11 bcl-1 locus in centrocytic lymphoma: analysis with multiple breakpoint probes. *Blood* **78:**493–498, 1991.
78. Williams, M. E., Swerdlow, S. H., and Meeker, T. C. Chromosome t(11;14)(q13;q32) breakpoints in centrocytic lymphoma are highly localized at the bcl-1 major translocation cluster. *Leukemia* **7:**1437–1440, 1993.
79. Campo, E., Raffeld, M., and Jaffe, E. S. Mantle cell lymphoma. *Semin. Hematol.* **36(2):**115–127, 1999.
80. Swerdlow, S. H. and Williams, M. E. From centrocytic to mantle cell lymphoma: a clinicopathologic and molecular review of 3 decades. *Hum. Pathol.* **33(1):**7–20, 2002.
81. Hanken, R. C. and Hunter S. V. Mantle cell lymphoma: molecular and immunophenotypic diagnostic aids. *Arch. Pathol. Lab. Med.* **123:**1182–1188, 1999.
82. Baron, B. W., Stanger, R. R., Hume, E., et al. Bcl-6 encodes a sequence specific DNA binding protein. *Genes Chromosomes Cancer* **13:**221–224, 1995.
83. Kerckaert, J. P., Deweindt, C., Tilly, H., et al. LAZ3, a novel zinc-finger encoding gene, is disrupted by recurring chromosome 3q27 translocations in human lymphomas. *Nat. Genet.* **5:**66–70, 1993.
84. Ye, B. H., Rao, P. H., Chaganti, R. S. K., and Dalla-Favera, R. Cloning of bcl-6, the locus involved in chromosome translocations affecting band 3q27 in B cell lymphoma. *Cancer Res.* **53:**2732–2735, 1993.
85. Kawamata, N., Nakamura, Y., Miki, T., et al. Detection of chimaeric transcripts of the immunoglobulin heavy chain and bcl6 genes by reverse transcriptase polymerase chain reaction in B cell non-Hodgkin's lymphomas. *Br. J. Haematol.* **100:**484–489, 1998.
86. Muramatsu, M., Akasaka, T., Kadowaki, N., et al. Rearrangement of the bcl6 gene in B cell lymphoid neoplasms: comparison with lymphomas associated with bcl2 rearrangement. *Br. J. Haematol.* **93:**911–920, 1996.
87. Niu, H. The proto-oncogene BCL-6 in normal and malignant B cell development. *Hematol. Oncol.* **20(4):**155–66, 2002.
88. Cattoretti, G., Chang, C. C., Cechova , K., et al. BCL-6 protein is expressed in germinal-center B cells. *Blood* **86(1):**45–53, 1995.
89. Lo Coco, F., Ye, B. H., Lista, F., et al. Rearrangements of the BCL-6 gene in diffuse large cell non-Hodgkin's lymphoma. *Blood* **83(7):**1757–1759, 1994.
90. Mason, D. Y., Bastard, C., Romokh, R., et al. CD30-positive large cell lymphomas ("Ki-1 lymphoma") are associated with a chromosomal translocation involving 5q35. *Br. J. Haematol.* **74:**161, 1990.
91. Morris, S. W., Kirstein, M. N., Valentine, M. B., et al. Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science* **263:**1281, 1994.
92. Lamant, L., Meggetto, F., Al Saati, T., et al. High incidence of the t(2;5)(q35) translocation in anaplastic large cell lymphoma and its lack of detection in Hodgkin's disease. Comparison of cytogenetic analysis, reverse transcriptase polymerase chain reaction, and P-80 immunostaining. *Blood* **87:**284–291, 1996.
93. Jaffe, E. S. Anaplastic large cell lymphoma: the shifting sands of hematopathology. *Mod. Pathol.* **14(3):**219–228, 2001.
94. Kutok, J. L. and Aster, J. C. Molecular biology of anaplastic lymphoma kinase-positive anaplastic large-cell lymphoma. *J. Clin. Oncol.* **20(17):**3691–3702, 2002.
95. Stein, H., Foss, H. D., Durkop H., et al. CD30+ anaplastic large cell lymphoma: a review of its histopathologic, genetic and clinical features. *Blood* **96(12):**3681–3695, 2000.
96. Haluska, F. G., Brufsky, A. M., and Canellos, G. P. The cellular biology of the Reed-Sternberg cell. *Blood* **84:**1005–1019, 1994.
97. Chan, W. C. Cellular origin of nodular lymphocyte-predominant Hodgkin's lymphoma: immunophenotypic and molecular studies. *Semin. Hematol.* **36(3):**242–252, 1999.

98. Stein, H. and Hummel, M. Cellular origin and clonality of classic Hodgkin's lymphoma: immunophenotypic and molecular studies. *Semin. Hematol.* **36(3):**233–241, 1999.
99. Marafioti, T., Hummel, M., Foss H. D., et al. Hodgkin and Reed–Sternberg cells represent an expansion of a single clone originating from a germinal center B-cell with functional immunoglobulin gene rearrangements but defective immunoglobulin transcription. *Blood* **95(4):**1443–1450, 2000.
100. Weiss, L. M., Strickler, J. G., Hu, E., Warnke, R. A., and Sklar, J. Immunoglobulin gene rearrangements in Hodgkin's disease. *Hum. Pathol.* **17:**1009–1014, 1986.
101. Marafioti, T., Anagnostopoulos, I., Foss, H. D., et al. Origin of nodular lymphocyte-predominant Hodgkin's disease from a clonal expansion of highly mutated germinal-center B cells. *N. Engl. J. Med.* **337(7):**453–458, 1997.
102. Corradini, P., Ladetto, M., Pileri, A., and Tarella, C. Clinical relevance of minimal residual disease monitoring in non-Hodgkin's lymphoma: a critical reappraisal of molecular strategies. *Leukemia* **13:**1691–1695, 1999.
103. Negrin, R. S. and Blume, K. G. The use of the polymerase chain reaction for the detection of minimal residual malignant disease. *Blood* **78:**255–258, 1991.
104. Sklar, J. Polymerase chain reaction: The molecular microscope of residual disease. *J. Clin. Oncol.* **9:**1521–1524, 1991.
105. Corradini, P., Stolfi, M., Cherasco, C., et al. Molecular monitoring of minimal residual disease in follicular and mantle cell non-Hodgkin's lymphomas treated with high-dose chemotherapy and peripheral blood progenitor cell autografting. *Blood* **89:**724–731, 1997.
106. Wu, G. Q., Sharp, J. G., Wu, G., et al. The detection of minimal lymphoma by molecular and combined culture-molecular methods. *Br. J. Haematol.* **99:**873–881, 1997.
107. Dolken, G. Detection of minimal residual disease. *Adv. Cancer Res.* **82:**133–185, 2001.

33 Molecular Genetic Abnormalities in Acute and Chronic Leukemias

Pei Lin and L. Jeffrey Medeiros

1. INTRODUCTION

Leukemias are monoclonal neoplasms that arise as a result of acquired mutations in hematopoietic progenitor cells. Approximately 25,000 to 30,000 new cases of leukemia are diagnosed in the United States each year *(1)*. Clinically, leukemias can be divided into acute and chronic neoplasms. Patients with acute leukemia usually present with an abrupt onset of symptoms attributable to bone marrow failure. Thus, patients can present with anemia, leukopenia, or thrombocytopenia. The neoplastic cells are immature blasts and the disease is aggressive, requiring immediate and aggressive chemotherapy. By contrast, patients with chronic leukemia have more gradual onset of symptoms, usually without bone marrow failure initially, and the neoplastic cells are differentiated (e.g., maturing granulocytes or small lymphocytes). Therapy for chronic leukemias is typically less aggressive at onset and is not needed for some patients (e.g., chronic lymphocytic leukemia), although eventually patients usually require therapy for either massive tumor burden or transformation to a higher-grade tumor.

Traditionally, morphologic, cytochemical, and immunophenotypic methods have been used to divide acute and chronic leukemias according to myeloid or lymphoid lineage. Of these four broad groups, acute myeloid leukemia (AML) is most common, approx 10,000 new cases diagnosed per year (Fig. 1), followed by chronic lymphocytic leukemia (CLL), with 8000 new cases per year, chronic myelogenous leukemia (CML), with 4500 new cases per year, and acute lymphoblastic leukemia (ALL), with 3100 new cases per year *(1)*.

Two classification systems, the popular French–American–British (FAB) and the more recent World Health Organization (WHO), are currently used to classify leukemias *(2,3)*. The recent WHO system, unlike the older FAB system, has incorporated into the classification more information regarding immunophenotypic and molecular abnormalities involved in the pathogenesis of leukemias. In these systems, myeloid leukemias can be subdivided into those of granulocytic, monocytic, erythroid, and megakaryocytic lineage and lymphoid leukemias can be subdivided into those of B-cell or T-cell lineage.

Although these leukemia classifications systems are extremely useful for the discussion of these diseases, in fact they do not incorporate the full extent of molecular information that is germane to the pathogenesis and/or treatment of leukemias. Knowledge of the molecular aspects of leukemias is also expanding almost daily. As a result, both the FAB and WHO classifications, by necessity, have oversimplified leukemias. For example, conventional cytogenetic and fluorescence *in situ* hybridization (FISH) analysis of AML cases have shown amazing heterogeneity and complexity, indicating that the category of AML in the FAB and WHO classifications is very heterogeneous. Using molecular methods, many of the genes involved in recurrent cytogenetic abnormalities have been cloned. Furthermore, to date, only a subset of all cases of AML are well characterized at the molecular level. Many other genes are likely to be discovered to be involved in subsets of AML. Many other categories of leukemia are also very heterogeneous at the molecular level, with CML being a prominent exception.

The discovery of these molecular abnormalities and recognition of the role of many of these abnormalities in the pathogenesis of leukemias has greatly improved our understanding of these diseases and our ability to diagnose them. As a result, some types or subsets of leukemia are currently defined by their molecular abnormalities. For example, *pml-rarα* is characteristic of a distinct type of AML, acute promyelocytic leukemia, with distinctive clinical and morphologic features that can be treated with all-*trans* retinoic acid (ATRA), followed by more traditional chemotherapy. Similarly, *bcr-abl* is now considered to be the *sine qua non* of CML, and a drug, imatinib mesylate (STI571), has been designed to specifically inhibit the tyrosine kinase activity of BCR-ABL. It is expected that leukemias will be further defined at the molecular level and that additional therapies will be directed at specific molecular abnormalities or pathways in the future.

Progress is this regard is likely to be facilitated by the use of gene expression profiling. Initial studies using this technology have shown that gene expression patterns are characteristic for many of the leukemias with distinctive molecular abnormalities

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

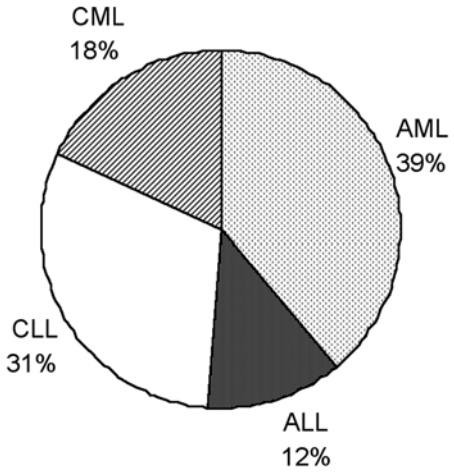


Fig. 1. Relative frequency of four broad disease groups discussed in this chapter and based on 1999 statistics is shown *(1)*. AML, acute myeloid leukemia; CML, chronic myelogenous leukemia; ALL, acute lymphoblastic leukemia (ALL); CLL, chronic lymphocytic leukemia.

(class prediction). In addition, these studies have identified the expression of previously unknown genes in these neoplasms, as well as patterns of gene expression suggestive of novel types of leukemia (class discovery) *(4,5)*.

In this chapter, we discuss the common and/or well-characterized molecular abnormalities in acute and chronic leukemias of myeloid and lymphoid lineage.

2. ACUTE MYELOID LEUKEMIAS

2.1. GENE MUTATIONS Gene mutations can be divided into those that have a dominant or recessive effect. In dominant gene mutations, only one gene copy (allele) is required for activation (e.g., classical oncogene). By contrast, in recessive gene mutations, both gene alleles need to be inactivated for effect (e.g., classical tumor suppressor gene).

2.1.1. Ras The *ras* gene superfamily encodes for approximately 50 structurally related proteins that range from 21 to 29 kDa. Each of these proteins has a consensus guanosine triphosphate (GTP)-binding motif that is important for intracellular signaling, hence the name G-protein. Only some of these family members have been studied in hematologic neoplasms, most commonly *n-ras* at chromosome 1p11-13, *k-ras* at chromosome 12p11-12, and *h-ras* at chromosome 11p15. *N-ras* was first identified in a neuroblastoma cell line. *K-ras* and *h-ras* are cellular oncogenes evolutionarily related to the Kirsten and Harvey murine sarcoma viruses, respectively. Mutations of *ras* genes are most commonly point mutations that result in gene activation.

Many types of myeloid leukemia can have *ras* gene mutations, with *n-ras* being involved most often, followed by *k-ras*. *H-ras* is rarely mutated in myeloid neoplasms. *Ras* mutations are most common in chronic myelomonocytic leukemia, reported in up to 40–50% of cases, followed by AMLs (30–40%), aggressive myelodysplastic syndromes (10–20%) and ALL (10–20%) *(6–8)*. *Ras* mutations are rare in CML *(8)*.

It seems likely that *ras* gene activation is underestimated in the literature for at least two reasons. First, in many mutational studies, only "hot spots" (codons 12, 13, and 61) have been assessed for mutations. Codons 12 and 13 occur in the first exon of the gene. Codon 61 is located in the gene promoter. Obviously, mutations also can occur, albeit at lower frequency elsewhere within the gene. Second, *ras* can be activated without mutation (e.g., *bcr-abl* activates *ras) (8)*.

2.1.2. Runx1/aml1 The *runx1* gene, also known as *aml1, cbfα2,* and *pebp2αB*, is located on chromosome 21q22. The *runx1* gene encodes core binding factor α, one of the two units that encode core binding factor and is involved in certain types of acute leukemia (discussed subsequently).

The *runx1* gene is mutated in a subset of myeloid neoplasms, approx 20%, most often in AMLs, but also in myelodysplastic syndromes *(9)*. The *runx1* gene is also amplified in a small subset of AML and ALL cases.

2.1.3. Rb The retinoblastoma (*rb*) gene, located at chromosome 13q14, is a classic tumor suppressor gene. Thus, both alleles need to be inactivated for loss of normal function. The *rb* gene encodes for a 110-kDa nuclear phosphoprotein that is intricately involved in the cell cycle. *Rb* gene abnormalities have been identified in approx 10% of AMLs, as well as in CML cases at time of blast crisis and aggressive myelodysplastic syndromes *(10)*.

2.1.4. p53 The *p53* gene, located at chromosome 17p11.2, is another classic tumor suppressor gene. The *p53* gene is one of the most commonly affected genes in cancers, although its frequency of involvement is myeloid leukemias is less than that of other genes such as *flt*3 or *ras*. *p53* inactivation occurs predominantly through a point mutation in one allele and deletion of the other allele *(11)*. The product of the *p53* gene is a nuclear phosphoprotein of 53 kDa that functions as a transcription factor. It also inhibits DNA replication, controls progression of the cell cycle, and is involved in the regulation of apoptosis.

p53 point mutations are found in approx 6% of AML cases, including all morphologic subtypes except acute promyelocytic leukemia (FAB-M3). *p53* gene mutations correlate with older patient age, presence of myelodysplasia, and poor prognosis *(6,11)*.

2.1.5. c-kit The *c-kit* gene, located on chromosome 4q12, encodes a 145-kDa transmembrane protein that is a class III receptor kinase. *C-kit* is normally expressed by myeloid, erythroid, and megakaryocytic precursors and mature mast cells. Mutations in the *c-kit* gene occur in a substantial subset of cases of AML with inv(16)(p13q22) or AML with t(8;21)(q22;q22). In a recent study, *c-kit* gene mutations involving exon 8 or Asp816 were detected in 20 of 63 (32%) of AML with inv(16) and 6 of 47 (13%) of AML with t(8;21) *(12)*. *C-kit* gene mutations are rare in other types of *de novo* AML.

2.2. PARTIAL TANDEM DUPLICATION OF GENES

2.2.1. Flt3 The *flt3* (*fms*-like tyrosine kinase 3) gene, also known as *flk2* (fetal liver kinase 2), is located on chromosome 13q12 and encodes a membrane-bound class III receptor tyrosine kinase *(13)*. This gene, which has 24 exons, plays an important role in normal hematopoiesis and has extensive homology with other class III receptor tyrosine kinases, including *c-kit* and platelet-derived growth factor (*pdgf*).

Flt3 gene mutations are one of the most common molecular abnormalities recognized in AMLs, identified in up to one-third

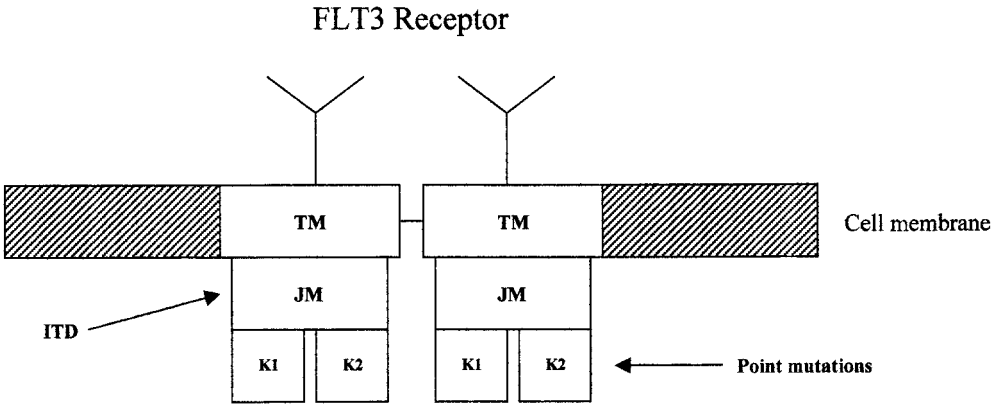


Fig. 2. FLT3 receptor. Internal tandem duplications (ITD) occur in exons 14 or 15 corresponding with the juxtamembrane domain (JM). Point mutations occur in exon 20 corresponding with the second tyrosine–kinase domain (K2). TM, transmembrane domain; K1, first tyrosine–kinase domain.

of cases in some studies *(14,15)*. These mutations activate the *flt3* tyrosine kinase, thereby playing a role in leukemogenesis. However, *flt3* gene mutations alone are insufficient to cause leukemia in transgenic mice models.

There are two types of *flt3* gene mutation (Fig. 2). The most common, by far, is an internal tandem duplication (ITD) that occurs between exons 14 and 15 (previously known as exons 11 and 12). These exons are located within the juxtamembrane domain of the gene. The ITD can be highly variable in length (3 to >400 bp). Nevertheless, this duplication is always inserted in a head-to-tail direction and is in-frame, thereby preserving the transcript. *Flt3* gene mutations of the ITD type have been detected in approx 25% of AMLs.

The less common type of *flt*3 gene mutation involves a single-basepair substitution (missense point mutation) in codon 835 of exon 20 (previously known as exon 17). Exon 20 is within the tyrosine–kinase domain of the gene. D835 mutations occur in 5–10% of AML cases. Whether or not ITD and D835 mutations activate the gene in a similar manner and effect downstream molecules similarly is unknown.

Flt3 gene mutations have been detected in many different types of AML, including most types designated by the FAB and WHO *(14,15)*. However, *flt3* gene mutations are not entirely specific for AMLs. A small subset of cases of myelodysplastic syndrome, 5–10%, carry *flt3* gene mutations (either type) and *flt3* gene mutations also have been detected in a small subset of myeloproliferative syndromes, usually in accelerated or blastic phase in our experience, and rarely in precursor B-cell ALL.

Both ITD and D835 types of *flt3* gene mutations are easily detected by polymerase chain reaction (PCR) assays. As ITD mutations alter product size and D835 mutations affect a restriction enzyme site, usually sequencing is not necessary.

2.2.2. mll The *mll* (mixed lineage leukemia) gene, also known as *all1*, *hrx*, and *htrx*, is well known for its involvement in a number of different chromosomal translocations (discussed subsequently). However, this gene also can be partially duplicated in tandem in cases of AML without evidence of translocations or other cytogenetic evidence of abnormalities involving 11q23 *(16)*. This process of duplication also has been referred to as *mll* self-fusion *(17)*.

These duplications have been associated with trisomy 11, but are not restricted to trisomy 11 cases *(18)*. *Mll* gene duplications occur within the area between exons 2 and 8 of the *mll* gene and are always in-frame. Thus, the duplicated regions are always transcribed. In all cases, the AT hook and DNA methyltransferase motifs are preserved, suggesting that these regions of the *mll* gene are critical for leukemogenesis. *Mll* partial tandem duplications can be detected by Southern blot analysis and PCR approaches.

2.3. CHROMOSOMAL TRANSLOCATIONS AND INVERSIONS A number of chromosomal rearrangements have been detected in AML. Some of these abnormalities are recurrent and correlate with distinctive clinical, morphologic, or immunophenotypic findings. In most instances, these rearrangements are reciprocal chromosomal translocations that disrupt two genes, recombining one end of each gene to form a novel chimeric gene. Chromosomal translocations usually affect only one allele and are dominant.

2.3.1. t(15;17)(q22;q21) This translocation is present exclusively in acute promyelocytic leukemia (APL), also known as AML-M3 in the FAB classification and as AML with t(15;17)(q22;q21) in the recent WHO classification *(2,3)*. Acute promyelocytic leukemia represents approx 5–8% of all cases of AML *(3)*.

Acute promyelocytic leukemia is perhaps the most distinctive type of AML. Patients commonly first present with coagulopathy and a bleeding dialysis, attributable to disseminated intravascular coagulation and primary fibrinolysis *(19)*. Morphologically, the neoplastic cells are distinctive promyelocytes of which there are two types, the granular and microgranular (hypogranular) variants. Promyelocytes of the common hypergranular variant (M3) have numerous azurophilic granules and Auer rods in their cytoplasm, often obscuring the nucleus. Promyelocytes of the microgranular variant (M3v) have many fewer cytoplasmic granules and often have bilobed ("apple-core") nuclei *(19,20)*. Immunophenotypic studies have also shown distinctive patterns of antigen expression, with the promyelocytes of M3 being commonly negative for CD34 and HLA-DR, in contrast with the promyelocytes of M3v *(20)*.

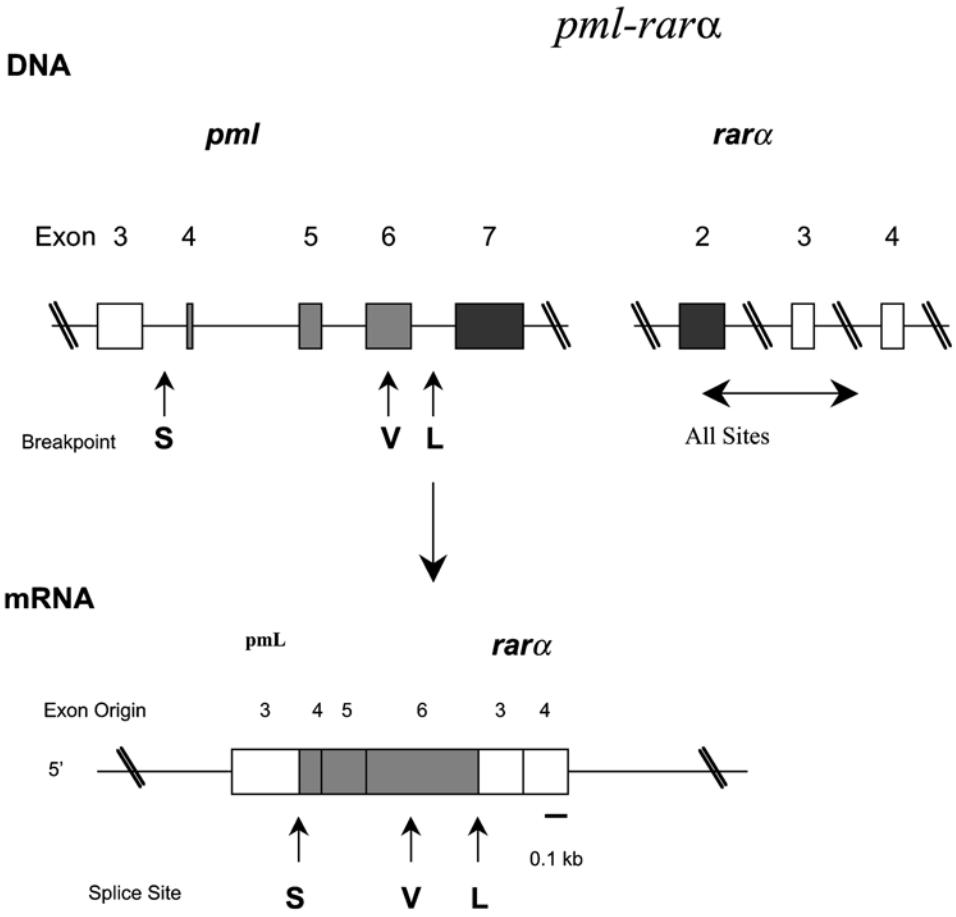


Fig. 3. t(15;17)(q22;q21). The upper figures illustrate the germline configuration of the *pml* and *rarα* genes. The *pml* and rarα genes reside on chromosomes 15 and 17, respectively. The *pml-rarα* fusion gene resides on the derivative chromosome 15. The lower figure shows the chimeric mRNA encoded by the fusion gene. The numbered boxes correspond to exons. Arrows show breakpoints that correspond to short (s), variable (v), and long *(l)* forms of *pml-rarα*.

In the t(15;17), the retinoic acid receptor alpha (*rarα*) gene on 17q21 is disrupted and fused with a nuclear regulatory factor, the promyelocytic leukemia (*pml)* gene on 15q22 giving rise to a *pml-rarα* gene fusion product on the derivative chromosome 15 (Fig. 3) *(21)*. The breakpoint in the *rarα* gene occurs consistently within intron 2. The breakpoints in the *pml* gene occur in three different sites resulting in three major forms of *pml-rarα*: intron 6 (bcr1, long form), intron 3 (bcr3, short form) and exon 6 (bcr2, variable form). The most frequent sites are bcr1 and bcr3 (each seen in 40–50% of cases). Some studies have suggested that the different forms of *pml-rarα* correlate with clinical presentation or prognosis, but the data are not definitive.

Wild-type *pml* has been shown to function as a tumor suppressor gene. PML in cells accumulates in discrete nuclear bodies and has been suggested to be involved in maintaining genome integrity *(22)*. *Rarα* has been shown to promote differentiation and suppress growth. The novel PML-RARα fusion protein is thought to play a key role in the pathogenesis of APL, by acting as a dominant negative molecule preventing transcription of genes targeted by RARα, thus interfering with myeloid maturation *(19)*. The mechanism of action of the novel protein, PML-RARα, is thought to involve binding with *N*-COR-histone deacetylase, thereby becoming a transcriptional repressor *(23)*.

Acute promyelocytic leukemia was the first hematologic neoplasm in which therapy successfully targeted the specific molecular abnormality, *pml-rarα*. All-*trans* retinoic acid (ATRA) targets RARα and arsenic, a relatively recent advance in therapy, targets PML. These therapies allow the leukemia cells to differentiate, greatly reducing the coagulopathy and bleeding diathesis. These therapies are often followed by more traditional chemotherapy regimens *(19)*.

2.3.1.1. Molecular Methods For Detection of pml-rarα

A number of methods can be employed to detect the t(15;17) *(21)*. Conventional cytogenetic methods detect the translocation in approx 90% of morphologically defined APL cases. Reasons for "false-negative" cytogenetics in APL include a subset of cases with the t(11;17) or other more rare variant translocations (discussed subsequently) and technical reasons, such as poor growth of the neoplastic clone and the existence of cryptic translocations or insertions that are difficult to recognize *(24)*. FISH analysis, because it allows assessment of interphase nuclei as well as dividing cells, is particularly useful in cases in which technical issues such as poor cell growth preclude adequate conventional cytogenetic analysis. These methods are also helpful in assessing minimal residual disease. Reverse transcriptase-polymerase chain reaction (RT-PCR) methods define the type of *pml* breakpoint and are the most

Partner	Chromosome	Chimeric Protein	Frequency	Response to ATRA
pml	15	PML-RARα	95%	Yes
plzf	11	PLZF- RARα	<5%	No
npm	5	NPM- RARα	<1%	Yes
numa	11	NUMA- RARα	<1%	Yes
stat5b	17	STAT5B-RARα	< 1%	No

Fig. 4. Various fusion genes reported in acute premyelocytic leukemia, with their frequency and response to all- *trans* retinoic acid (ATRA).

sensitive means of monitoring minimal residual disease. Virtually all cases of APL can be detected by restriction fragment length analysis (RFLA) using genomic *pml* or *rarα* probes. However, these methods are also time consuming and laborious and are not used routinely in most laboratories.

2.3.1.2. Immunofluorescence Methods for Detection of pml-rarα Assessment of the distribution of PML in leukemic cells provides a simple, rapid, low-cost approach for diagnosis *(25)*. In normal myeloid cells, PML has a speckled (or macrogranular) pattern of staining which corresponds to the localization of PML into 5–20 discrete nuclear bodies or PML oncogenic domains (PODs). In contrast, APL cells have a characteristic and easily distinguishable nuclear staining pattern known as "microspeckled," resulting from the disruption of the nuclear bodies and redistribution of PML into greater than 50 smaller granules per cell. The monoclonal antibody PG-M3, directed against an amino-terminal epitope of PML that is shared by all the known PML-RARα isoforms, is particularly suitable for diagnostic use *(25)*. In some institutions (including our own), immunofluorescence analysis for PML has been incorporated into the standard diagnostic work-up of patients with suspected APL.

2.3.2. Variant Translocations Associated with Acute Promyelocytic Leukemia Almost all (95%) cases of APL carry the t(15;17), although a small subset of these cases have nontypical translocations, resulting from insertions or three-way translocations (Fig. 4) *(24)*. The remaining cases morphologically resemble APL but carry variant translocations. At least four variant translocations involving *rarα* have been described to date. Most these neoplasms carry the t(11;17). There are also rare cases of APL in which the molecular lesion resulting in leukemia is not recognizable at this time *(24)*.

2.3.2.1. t(11;17)(q23;q21) Many patients with APL that carry the t(11;17)(q23;q21) present with bleeding diathesis and also clinically resemble classical APL *(26)*. Unlike APL, t(11;17)-positive AML responds poorly to ATRA treatment, although it may respond to ATRA combined with granulocyte colony-stimulating factor *(27)*.

The t(11;17) is a reciprocal and balanced translocation involving the *rarα* gene on chromosome 17q21 and the promyelocytic leukemia zinc-finger (*plzf*) gene on chromosome 11q13 *(28,29)*. Two fusion genes are produced as a result of the translocation, *plzf-rarα* and *rarα-plzf*. It is presently uncertain which chimeric protein is oncogenic. The structure of both chimeric proteins suggests that either can bind to DNA and influence transcription. *plzf* is a zinc-finger transcription factor and its protein product is thought to negatively regulate the cell cycle, ultimately leading to growth suppression. RARα-PLZF activates *cyclin* A2 transcription that inhibits growth suppression, which might be the oncogenic mechanism of this fusion protein *(30,31)*. In t(11;17), the breakpoint in the *rarα* gene occurs in intron 2, similar to the t(15;17). The breakpoints within *plzf* are variable.

2.3.2.2. t(11;17)(q13;q21) Rare cases of APL with the t(11;17)(q13;q21) are reported *(24)*. The t(11;17) is a reciprocal translocation involving the nuclear mitotic apparatus (*numa)* gene on chromosome 11 and the *rarα* gene on chromosome 17. These neoplasms are responsive to ATRA.

2.3.2.3. t(5;17)(q23;q21) Similarly, rare cases of APL with the t(5;17) are reported *(24)*. The t(5;17) is a reciprocal translocation involving the *npm* (nucleophosmin) gene on chromosome 5 and the *rarα* gene on chromosome 17. These neoplasms are responsive to ATRA.

2.3.2.4. t(17;17)(q11;q21) One report describes a case of APL with the t(17;17) involving the *stat5b* gene at chromosome 17q11 and the *rarα* gene on chromosome 17q21 *(32)*. These neoplasms are not responsive to ATRA.

2.3.3. t(8;21)(q22;q22) The t(8;21)(q22;q22) is found in 5–12% of cases of AML *(3)*. The t(8;21) is one of the core binding factor (CBF) leukemias, along with the inv(16) and t(3;21) in AML and the t(12;21) in pediatric ALL. The CBFs are part of a family of heterodimeric transcription factors containing a common β-subunit, CBFβ, and one of three CBFα subunits.

t(8;21)-positive AML has distinctive morphologic and immunophenotypic features *(28)*. Morphologically, most of these leukemias are of M2 type as defined by the FAB classification. More than 90% of cases of AML with t(8;21) have M2 morphology, and as many as 30–40% of M2 cases have t(8;21). In addition, the blasts in t(8;21)-positive AML often have long, thin Auer rods and salmon-colored cytoplasmic granules. Immunophenotypic studies have shown that these neoplasms can express CD56 (associated with poorer prognosis) and B-cell antigens, such as CD19 and BSAP/PAX5 *(28)*. The incidence of t(8;21)-positive leukemias is higher in younger patients, and most patients with *de novo* AML associated with t(8;21) respond well to chemotherapy *(29)*.

The t(8;21) is a reciprocal chromosomal translocation involving the *eto* gene (also known as *mtg8*) on chromosome 8q22 and the *runx1* (runt related transcription factor 1) gene (often referred to as *aml1* in the literature) on chromosome 21q22 *(33,34),* with the *runx1-eto* fusion gene residing on the derivative chromosome 8. The *runx1* gene encodes the CBFα protein that normally binds to the DNA core-enhancer sequence, TGTGGT, and also binds with CBFβ. The runt-homology domain of RUNX1 is required for DNA binding and heterodimerization (Fig. 5). CBFβ does not bind to DNA directly, but via its binding to RUNX1, it improves the efficiency of RUNX1 binding to DNA. The RUNX1/CBFβ

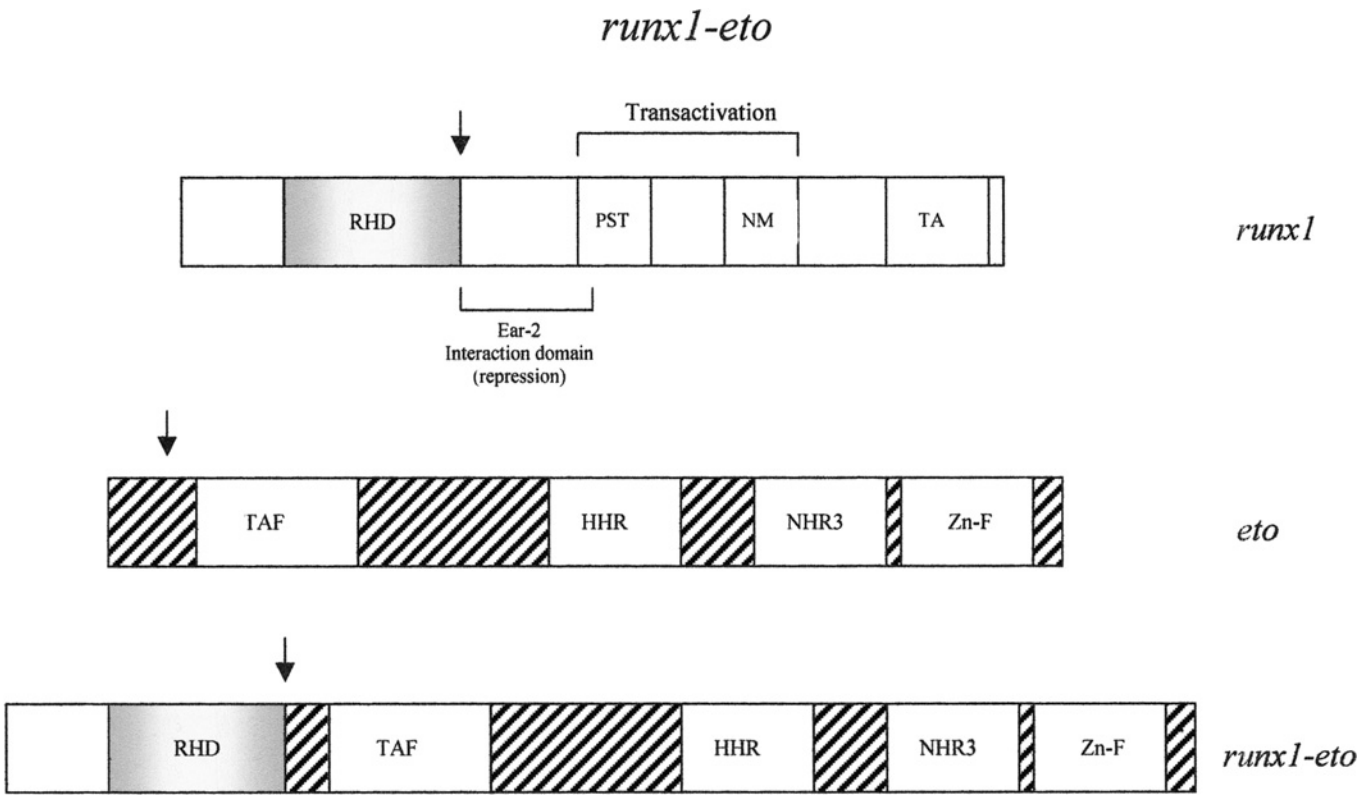


Fig. 5. Normal *runx* and *eto* proteins and the *runx1-eto* fusion protein encoded as a result of the t(8;21)(q22;q22). Arrows indicate breakpoints. RHD, runt homology DNA-binding domain; PST, proline/serine/threonine-rich region; NM, nuclear matrix attachment sequence; TA, transcriptional activation domain; TAF, domain with homology to transcription-activating factors; HHR, hydrophobic heptad repeat; NHR3, Nervy homology region 3; Zn-F; zinc-finger motifs.

complex then recruits transcription factors and coregulators (Fig. 6). The *eto* gene encodes the mammalian homolog of the *Drosophila* protein Nervy. In *Drosophila*, Nervy normally functions to repress transcription. Thus, RUNX1-ETO is thought to dominantly repress normal transcriptional activity of RUNX1.

Although the t(8;21) is a reciprocal translocation in most cases, rare AML cases with complex three-way (8;21;variable) translocations have revealed that the der(8) chromosome encodes the RUNX1-ETO fusion protein, which is most important in oncogenesis. The chromosomal breakpoints resulting from t(8;21) cluster within a single intron between exons 5 and 6 of *runx1* and within an intron at the 5′ region of *eto*. The encoded fusion protein consists of the N-terminal 177 amino acids of RUNX1 fused in frame to nearly full-length ETO protein *(33)*.

Studies in transgenic mice have shown that *runx1-eto* is insufficient for AML to develop *(35,36)*. In fact, RUNX1-ETO-positive mice have only minimal hematopoietic abnormalities. However, if other mutations are induced, these mice rapidly develop overt AML, including extramedullary granulocytic sarcomas. Thus, the t(8;21) is insufficient but necessary for leukemogenesis to occur.

In addition to its role as a transcriptional repressor, RUNX1-ETO also activates transcription of the BCL-2 promoter through an *runx1*-binding site that resides within a negative regulatory region of the promoter *(37)*. RUNX1-ETO can also bind to and inhibit $p14^{ARF}$ involved in the MDM2/p53 pathway *(38)* and has been shown to bind with PLZF *(39)*. Fusion of the ETO protein with RUNX1 also causes misdirection of RUNX1 to subnuclear foci distinct from those containing the wild-type protein *(40)*.

2.3.3.1. Molecular Methods For Detection of t(8;21) Conventional cytogenetics is most commonly used at time of initial diagnosis and detects the t(8;21) in over 95% of cases *(41)*. These methods are less suitable for assessing minimal residual disease because of lack of sensitivity. For the latter, RT-PCR methods are very sensitive and can detect almost all cases with the t(8;21), including 5–10% of *runx1-eto*-positive AML cases not detected by conventional cytogenetics.

2.3.4. Inv(16)(p13q22) or t(16;16)(p13;q22) The WHO classification recognizes AML with inv(16)(p13q22) or the much less frequent AML with t(16;16)(p13;q22) as a distinct entity *(3)*. In the FAB system, these neoplasms are known as M4Eo *(2)*. AML with inv(16) represents 10–12% of all cases of AML and has distinctive morphologic features. Most cases are myelomonocytic and are associated with dysplasia and abnormal eosinophils with prominent basophilic granules in the peripheral blood and bone marrow *(3,42,43)*. These eosinophils have been shown to be a part of the neoplastic clone *(44)*.

Both the inv(16)(p13q22) and the t(16;16)(p13;q22) result in a chimeric fusion protein composed of the amino terminus of the core binding factor β (*cbfβ*) gene at 16q22 with the smooth muscle myosin heavy chain (*myh11*) at 16p13 (Figs. 7 and 8) *(45–47)*. In both inv(16) and the t(16;16), the breakpoints in the *cbfβ* gene are similar, preserving the coding region. In the *myh11* gene, the breakpoints are variable and alternative

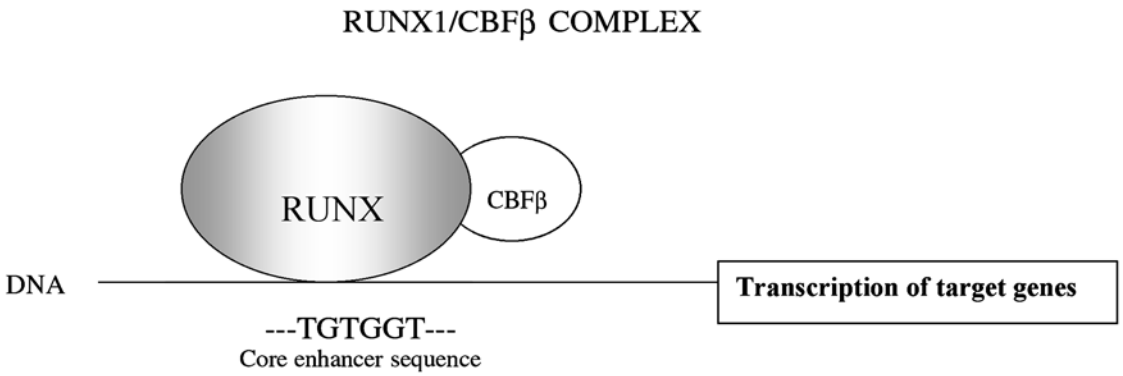


Fig. 6. Simplified schematic of RUNX/CBFβ complex. This complex binds to the DNA core-enhancer sequence and functions to organize other factors and induce transcription of target genes.

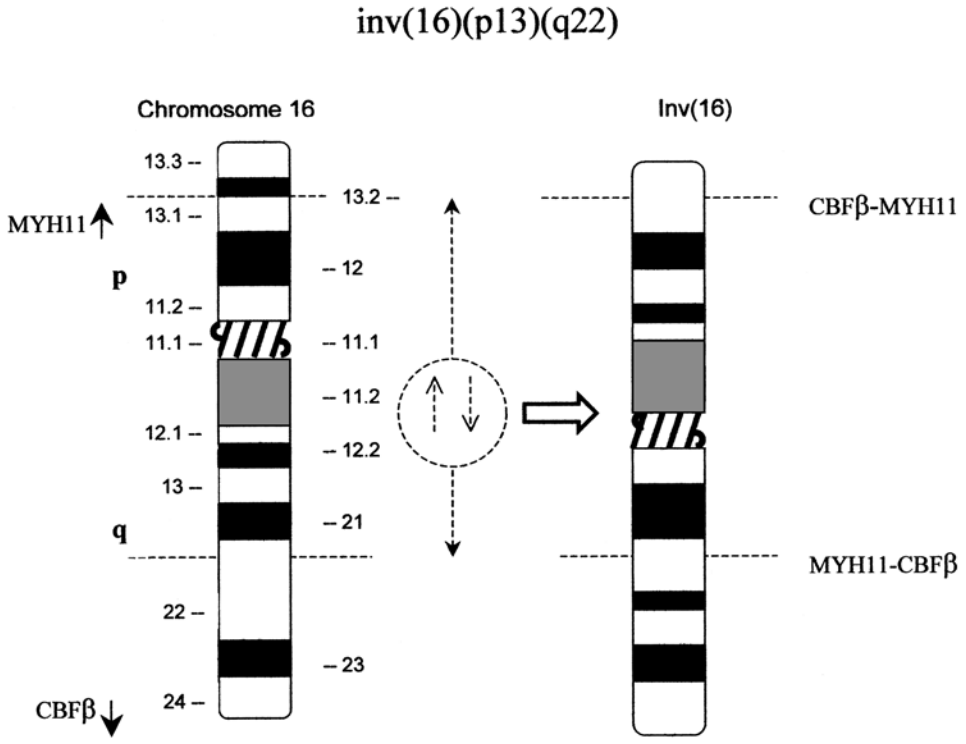


Fig. 7. Idiogram of the inv(16)(p13q22). The inv(16) is, by far, the most common molecular abnormality associated with AML with abnormal bone marrow eosinophils (FAB-M4Eo).

splicing also occurs. Thus, the number of *cbfβ-myh11* transcripts is at least eight. However, one type (so-called type A) of *cbfβ-myh11* occurs in more than 90% of cases.

CBFβ does not bind DNA directly, but increases the affinity of the CBFα subunits for binding to DNA *(47)*. CBFβ is critical to normal hematopoiesis, as knockout mice fail to develop definitive hematopoiesis. One hypothesis is that CBFβ-MYH11 sequesters CBFα in the cytoplasm as nonfunctional complexes. This complex in the cytoplasm also recruits other molecules (e.g., histone deacetylase), thereby acting as a transcriptional repressor *(47)*. However, the occurrence *cbfβ-myh11* alone is insufficient to cause AML in transgenic mice; additional mutations are also required for AML to develop.

2.3.4.1. Molecular Methods For Detection of inv(16) The conventional cytogenetics approach is most commonly used at time of initial diagnosis and detects inv(16) or t(16;16) in most cases. However, inv(16) can be subtle and therefore missed in a small subset of cases. RT-PCR assays for precise diagnosis of AML with the inv(16) or the t(16;16) has been advocated. Up to 10% of AML cases with detectable *cbfβ-myh*11 transcripts can be negative by conventional cytogenetic analysis. In some instances, RT-PCR detection might lead to retrospective detection of the cytogenetic abnormality. FISH also has been used successfully to detect the inv(16) *(48)*.

An immunofluorescence assay using a polyclonal antibody directed against the most common type of fusion protein (type A) has recently been produced and shown to have 100% concordance with conventional cytogenetics. However, this antibody is specific for the type A fusion protein and does not detect other transcript types *(49)*.

2.3.5. *mll* Gene Translocations The *mll* (myeloid/lymphoid leukemia or mixed lineage leukemia) gene (also known as

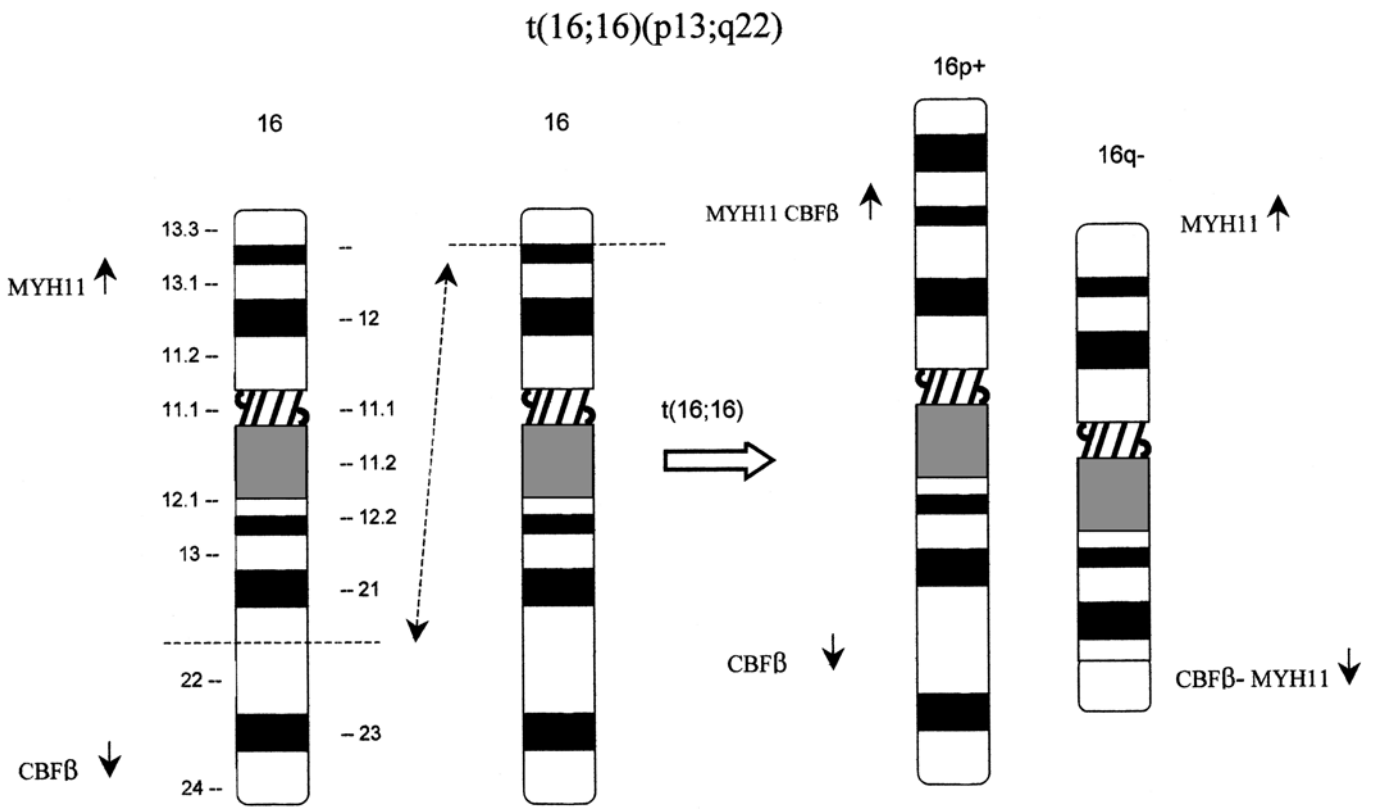


Fig. 8. t(16;16)(p13;q22). Idiogram of the t(16;16), a rare abnormality associated with AML with abnormal bone marrow eosinophils (FAB-M4Eo).

all-1, hrx, htrx-1) located at chromosome 11q23 is involved in recurrent translocations in 5–6% of AMLs according to the WHO classification *(3)* and in 2.8% of AMLs in a large German study *(50)*. The *mll* gene also can be amplified in a subset of AMLs *(51)*. Translocations and amplification of *mll* can be detected in many FAB subtypes of AML, most often in M4 or M5, but rarely (if ever) in M3, M6, and M7 *(50,51)*. Approximately 20–30% of the AML-M5, are associated with 11q23 translocations, the most common being t(9;11)(p22;q23). 11q23 translocations involving the *mll* gene are also involved in therapy-related leukemias that develop after exposure to topoisomerase II inhibitors *(3,50)*.

The *mll* gene is composed of 21 exons that span approx 100 kb and encodes a protein of 430 kDa. This gene is highly promiscuous with up to 50 different partner chromosomes that can participate in 11q23 translocations. The most common partner genes involved in these translocations are located at 6q27, 9p22, 10p12, 17q21, or 19p13.1 *(50,52)*. Approximately 25 of the partner genes have been cloned, and these partners encode for proteins involved in DNA binding, transcriptional activation, and chromatin remodeling. As a result of translocation, fusion proteins are formed that consist of the 5′ portion of the *mll* gene, fused to the 3′ portion of a gene encoded on the reciprocal chromosome and located on the derivative chromosome 11. The structure of *mll* suggests that its normal function is most likely that of DNA binding and chromatin modulation *(51)*.

Translocations involving the *mll* gene in acute leukemias, regardless of lineage, have been recently shown to upregulate *hox* gene expression *(53)*. The *hox* gene family includes at least 21 genes, but the genes overexpressed in these leukemias are *hoxa4*, *hoxa5*, *hoxa9*, *hoxa10*, and *hoxc6*. Of these, *hoxa9* and *hoxa10* are most highly expressed.

2.3.5.1. Molecular Methods For Detection of mll *Gene Translocations*

Conventional cytogenetics is the most complete approach for detecting 11q23 abnormalities, allowing the detection of translocations involving 11q23 and all partner chromosomes. However, a small subset of cases can be missed for both technical and interpretative reasons. Using Southern blot analysis or FISH methods, a single *mll* probe can detect most of the common and uncommon 11q23 breakpoints. However, these methods do not allow detection of the partner chromosomes without prior knowledge of the partner (usually obtained by conventional cytogenetic analysis). Because of the large number of possible partners, PCR methods are not practical for detecting 11q23 translocations routinely. With a prior knowledge of the partner, these methods can be used to detect minimal residual disease. Multiplex RT-PCR techniques using a number of primers have been developed to detect a number of the more common 11q23 translocations *(54)*.

2.3.6. t(6;9)(p23;q34)

The t(6;9)(p23;q34) has been identified in approx 1% or less of AML cases *(3)*. Using the WHO classification, with the cutoff of 20% blasts for AML, most t(6;9)-positive neoplasms are classified as AML *(3)*. However, using the FAB system, a subset of cases are classified as aggressive myelodysplastic syndromes *(2)*.

Patients with t(6;9)-positive AML are commonly of younger age (less than 40 yr) and have a poor prognosis *(55,56)*. Most cases are associated with myelodysplasia in the background and basophilia is common, in approximately half of these cases. Although clinical remission can be achieved with chemotherapy in a large subset of patients, relapse is common *(55,56)*.

The t(6;9) disrupts the *dek* gene at chromosome 6p23 and the *can* (*NUP214*) gene at chromosome 9q34, resulting in a

dek-can fusion gene on the derivative 6 chromosome that encodes a chimeric mRNA and protein *(55,57)*. The breakpoints in the t(6;9) are consistent and clustered. The breakpoints in the *dek* gene occur in one intron on chromosome 6, known as *icb-6* (for intron-containing breakpoint). The breakpoints in the *can* gene also cluster in one intron, known as *icb-9*. The *dek* gene is approx 40 kb in size and encodes a 43-kDa protein, located in the cell cytoplasm and thought to be a transcription factor. The *can* gene is relatively larger (over 140 kb) and encodes a 214-kDa protein. The normal CAN protein is a component of the nuclear–pore complex (hence, it is also known as nucleoporin) involved in the transport of mRNA and proteins between the cytoplasm and the nucleus *(58)*. The *dek-can* fusion protein, of predicted size of 165 kDa, has a nuclear distribution, suggesting that it is a transcription factor *(59)*.

*Flt*3 gene mutations are particularly commonly in t(6;9)-positive AML cases. Recent large studies of AMLs screened for *flt3* gene mutations have shown *flt3* gene mutations in over half of these neoplasms *(14,15)*. In our own experience, seven of eight cases assessed at our institution had *flt3* gene mutations, all of the internal tandem duplication type *(56)*.

2.3.6.1. MOLECULAR METHODS FOR DETECTION OF THE t(6;9) The clustering of breakpoints in icb-6 of *dek* and icb-9 of *can* allows convenient detection by using RT-PCR methods *(55)*. Conventional cytogenetics detects the t(6;9) and allows detection of additional abnormalities that may occur with disease progression; trisomies of 8 and 13 are the most common additional abnormalities *(55)*. Southern blot analysis with genomic probes derived from the *dek* and *can* genes also can be used.

2.4. ACUTE MEGAKARYOBLASTIC LEUKEMIA Acute megakaryoblastic leukemia, also known as AML-M7 in the FAB classification *(2),* is uncommon and represents less than 5% of all AML cases *(3)*. As AML-M7 is currently defined, this category is heterogeneous. As these neoplasms are often associated with myelofibrosis, precluding bone marrow aspiration and limiting material for molecular studies, AML-M7 is poorly understood at the molecular level. However, there are two small subsets of pediatric AML-M7 in which molecular information has become available recently.

2.4.1. Down's Syndrome Patients with Down's syndrome are genetically predisposed to AML, most commonly the AML-M7 type *(3)*. Affected patients usually develop AML-M7 at around 2 yr of age. Usually within days after birth, Down's syndrome patients also can develop a transient myeloproliferative disorder (TMD), characterized by leukocytosis and blasts morphologically indistinguishable from AML-M7 blasts. These lesions usually spontaneously regress.

Mutations of the *gata-1* gene, located on chromosome Xp11.23, occur very commonly in, and are restricted to, both TMD and AML-M7 in Down's syndrome patients (Fig. 9) *(60)*. *Gata-1* gene mutations are not reported in normal cells or other types of AML occurring in this patient group. The *gata-1* gene is the founding member of a family transcription factors and encodes a 50-kDa zinc-finger transcription factor important in normal development of erythroid cells and megakaryocytes *(60)*.

Gata-1 gene mutations usually occur in exon 2, introducing a premature stop codon and resulting in a shorter protein of 40 kDa lacking its amino-terminal transactivation domain. This short form, also known as GATA-1S, is thought to be involved in leukemogenesis.

p53 gene mutations have been reported in a few cases of AML-M7, but not in TMD cases, suggesting that *gata-1* gene mutations are the first "hit," with *p53* gene mutations or other abnormalities being the second "hit" and resulting in AML-M7 *(60)*.

2.4.2. t(1;22)(p13;q13) The t(1;22) is an uncommon abnormality in AML-M7, with approx 50 cases reported in the literature. Two research groups have identified fusion genes in these neoplasms using FISH methods. Ma and colleagues *(61)* described a fusion gene involving the *rbm15* gene (RNA-binding motif protein-15) located on chromosome 1p13 and the *mkl1* gene (megakaryoblastic leukemia-1) located on chromosome 22. The *rbm15-mkl1* fusion gene resides on the derivative chromosome 22. Mercher and colleagues *(62)* described a fusion between the *ott* (one two two) and *mal* (megakaryocytic acute leukemia) gene on chromosome 22, with the *ott-mal* fusion located on the derivative chromosome 22.

3. CHRONIC MYELOGENOUS LEUKEMIA

Chronic myelogenous leukemia (CML) is a chronic myeloproliferative disorder of hematopoietic stem cells that carry the t(9;22)(q34;q11.2). Patients present with marked leukocytosis and basophilia in the peripheral blood and bone marrow and usually have splenomegaly. After an initial *chronic phase* of 3–5 yr, the disease invariably progresses to a more aggressive *accelerated phase* with increased blasts and terminates in the *blast phase* (acute leukemia) of myeloid or lymphoid lineage *(3,63)*.

Chronic myelogenous leukemia was the first human neoplasm found to be associated with a recurrent chromosomal abnormality, the minute 22 or Philadelphia chromosome (Ph), named after the city where it was first discovered. The Ph results from the t(9;22)(q34;q11.2) that fuses 3′ sequences from the proto-oncogene *c-abl*, the cellular homolog of the transforming sequence of Abelson murine leukemia virus (v-*abl*) located on chromosome 9, with 5′ sequences of *bcr* located on chromosome 22 (Fig. 10) *(3,63)*. The *bcr-abl* fusion gene is oncogenic and is present in granulocytic, erythroid, megakaryocytic, and B-lymphoid precursors, supporting the concept that CML arises from transformed pluripotent stem cells. Chronic myelogenous leukemia is now defined by the presence of t(9;22)(q34;q11.2) or its characteristic fusion gene *bcr-abl (63)*.

In 5–10% of CML patients, *bcr-abl* fusion is generated by variants of the standard Ph that include one or several other chromosomes in addition to chromosomes 9 and 22. Variants of Ph can be cryptic or masked requiring higher-resolution banding or molecular methods for their recognition *(64)*. No clinical or prognostic differences are known between CML cases with variant or standard Ph at time of diagnosis, and patients are treated similarly.

The *c-abl* gene spans 230 kb and contains 11 exons (from 5′ to 3′) numbered 1b, 1a, and a2 through a11 (Fig. 11) *(63)*. The normal *c-abl* gene product, $p145^{abl}$, contains a myristoylation site, regulatory *src*-homology domains SH3 and SH2, a tyrosine kinase domain, a proline-rich domain, nuclear localization signals, a DNA-binding domain, and an actin-binding domain (Fig. 12). $p145^{abl}$ encodes a tyrosine kinase that is probably involved in cell cycle regulation *(65)* and its expression level decreases with granulocytic differentiation *(66)*.

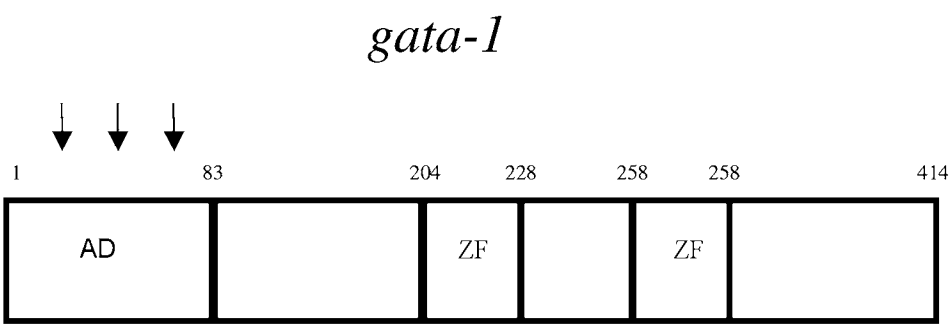


Fig. 9. *Gata*-1 gene located on chromosome Xp11.23. This gene is mutated in transient myeloproliferative disorder and acute megakaryoblastic leukemia of Down's syndrome. Arrows indicate the site of all described mutations, in the N-terminal activation domain (AD).

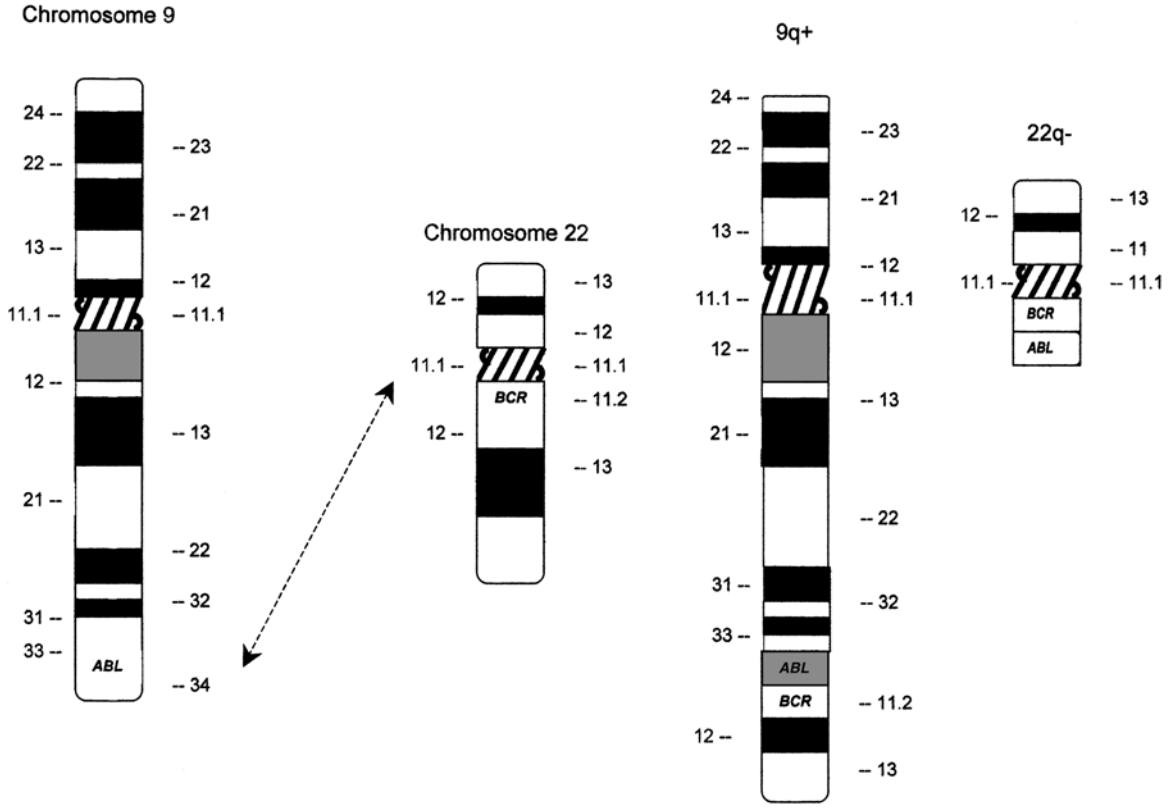


Fig. 10. Idiogram of the t(9;22)(q34;q11.2). The *bcr-abl* fusion gene resides on the derivative chromosome 22 (22q–).

The entire *bcr* gene consists of 21 exons that stretch over a region of 130 kb (Fig. 11) *(63)*. The *bcr* gene product encodes for a serine/threonine kinase, $p160^{bcr}$, which contains an oligomerization domain, a serine–threonine kinase domain (SH2) at the N-terminal, a central GEF domain, and a GAP (guanosine triphosphatase-activating protein) homology domain at the C-terminal (Fig. 12) *(63,67)*. $p160^{bcr}$ might be involved in signal transduction pathways and cytoskeletal/chromatin organization.

3.1. BCR-ABL *AND P210$^{bcr\text{-}abl}$* In CML, the breakpoints in *c-abl* occur over a 300-kb segment of the gene, either upstream of exon 1b, downstream of exon 1a, or between these two exons. The exons 1b and 1a are usually deleted in the final *bcr-abl* mRNA transcript, which generally contains exons a2–a11 of *c-abl (63)*. The breakpoints in *bcr* usually occur in the M-bcr, a 5.8-kb region that encompasses exons 12–16 (also known as exons b1–b5) (Fig. 11) *(3,63,67)*. In 95% of CML cases, the breakpoints arise between exons 13 (b2) and 14 (b3) or between exons 14 (b3) and 15 (b4). The *bcr-abl* fusion gene generated by the t(9;22)(q34;q11.2) is transcribed into two alternative mRNA products, e13a2 or e14a2 (also known as b2a2 and b3a2), that are approx 8.5 kb long. Occasionally, both transcripts are present in the same neoplasm. Translation of the mRNA transcripts yields two very similar p210 $^{bcr\text{-}abl}$ fusion proteins that give rise to CML.

The oncogenic ability of p210 $^{bcr\text{-}abl}$ is directly related to its aberrant tyrosine kinase activity, and it activates several signaling pathways through recruitment of adaptor proteins. These pathways include *ras*, STAT (signal transducer and activator of transcription), NFκβ, *c-jun*, phosphinositide-3-kinase (PI-3K)/AKT, reactive oxygene species, and MAPK (mitogen-activated protein kinase) *(63,67)*. The *ras* pathway appears to be pivotal for CML pathogenesis.

3.2. VARIANT *BCR-ABL* TRANSCRIPTS The m-*bcr* region occurs upstream of M-*bcr*, between exons e1 and e2,

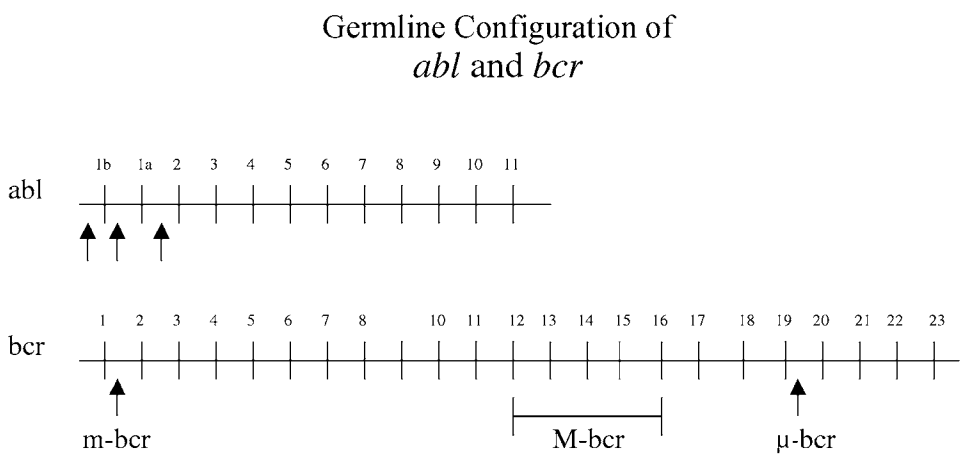


Fig. 11. Germline configuration of *abl* and *bcr* genes. Cross marks indicate exons and the arrows indicate breakpoints that result in various *bcr-abl* fusion genes.

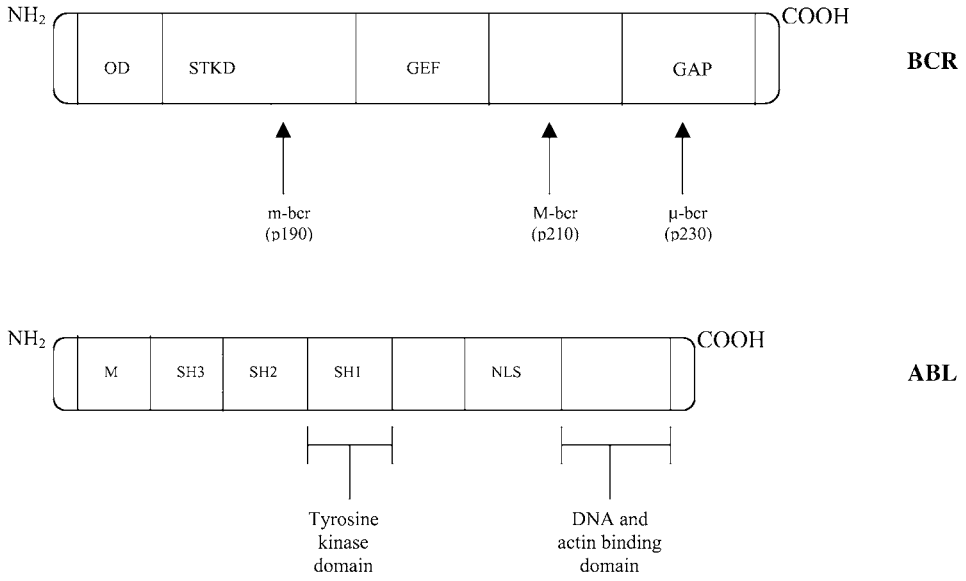


Fig. 12. Simplified schematic of BCR and ABL proteins. Arrows indicate breakpoints (as in Fig. 11). OD, oligodimerization domain; STKD, serine/threonine kinase domain; GEF, guanine exchange factor domain; GAP, guanosine triphosphatase-activating protein; M, myristoylation site; SH, *src*-homology domains; NLS, nuclear localizaton signal. See ref. *63* for additional details.

and produces an e1a2 transcript that encodes a smaller, $p190^{bcr\text{-}abl}$ fusion protein *(63,67)*. The m-*bcr* occurs in approx 25% of adult and in most pediatric cases of precursor B-lymphoblastic leukemia/lymphoma. Only rare cases of CML have breakpoints in the m-*bcr*, and these neoplasms tend to be associated with monocytosis and dysplasia. The μ-*bcr* occurs between exons e19 and e20 and encodes a larger fusion protein, $p230^{bcr\text{-}abl}$. The $p230^{bcr\text{-}abl}$ is associated with a rare form of CML that has prominent neutrophilic maturation *(68)*.

3.3. PH-NEGATIVE CML A small number of cases historically classified as Ph-negative CML share some features of CML, but lack cytogenetic evidence of the t(9;22)(q34;q11.2). Some of these cases are probably examples of CML carrying cryptic or masked translocations that can only be detected by more sensitive techniques such as FISH or RT-PCR. However, most of these cases have clinical behavior more akin to chronic myelomonocytic leukemia (CMML) and have molecular findings associated with CMML, including increased trisomy 8, *ras* gene mutations, and shorter overall survival *(3)*. These cases are best classified as a myeloproliferative/myelodysplastic disease *(3)*.

3.4. DELETIONS OF DERIVATIVE CHROMOSOME 9 The t(9;22)(q34;q11.2) is not always a balanced reciprocal translocation. Dual-color, dual-fusion FISH (D-FISH) methods have detected loss of portions of the reciprocal *abl-bcr* gene, on the derivative chromosome 9, in 10–15% of CML patients *(69)*. These deletions are primary events that occur at the time of translocation and are detected more frequently in neoplasms with a variant Ph. Chromosome 9 deletions usually span the translocation breakpoint and are large, with a size range reported in the literature from a few hundred kilobases to 8 mb.

Chronic myelogenous leukemia patients with chromosome 9 deletions are clinically indistinct at time of initial diagnosis, but tend to show early progression to advanced disease, with median survival times approximately half that of CML patients whose neoplasms do not carry these deletions. The critical genes that

are deleted are not entirely known. The argininosuccinate synthetase (*ass*) gene on chromosome 9 is one candidate *(69)*.

3.5. ACCELERATED OR BLAST PHASE The *accelerated* and *blast phases* of CML are frequently associated with secondary chromosomal changes, in approx 75% of all cases, including trisomy 8 (34%), additional Ph (30%), isochromosome 17 (20%), isochromosome 19 (13%), trisomy 21 (7%), trisomy 17 (5%), and monosomy 7 (5%) *(67)*. Rarely, translocations characteristically associated with distinct subtypes of *de novo* AML are detected at the time of the blast phase, including the t(8;21)(q22;22), t(3;21)(q26;q22), t(7;11)(p15;p15), inv16 (p13q22), and t(15;17)(q22;q12-21). Affected patients present with clinical, morphologic, and laboratory findings similar to patients with the corresponding *de novo* AML carrying these translocations.

At the molecular level, a variety of low-frequency acquired events have been reported in association with accelerated- and blast-phase CML, suggesting that maturation arrest might occur as a result of interruption of several pathways. Alteration of the *p53* gene is the most frequently found abnormality, occurring in 25% of cases of myeloid and megakaryocytic blast crisis *(70)*. Deletions of the $p16^{INK4A}$ gene can be detected in up to 50% of cases of lymphoid blast crisis *(71)*. Other less frequent genetic events include mutations of the *n-ras* and *rb* genes and mutation or overexpression of the *n-ras*, *c-myc,* or *evi1* genes.

3.6. MOLECULAR METHODS OF DETECTION OF bcr-abl Identification of the t(9;22)(q34;q11.2) by conventional cytogenetics is strong presumptive evidence of the *bcr-abl* fusion gene *(64)*. Conventional cytogenetics is usually performed on 20–25 metaphases and permits identification of the Ph as well as variant Ph and additional chromosomal changes. Techniques for identification of *bcr-abl* include FISH and RT-PCR. FISH is required to identify a cryptic Ph as well as deletions of derivative 9. Using FISH, up to 500 cells from peripheral blood or bone marrow aspirate are assessed, allowing quantification of the number of neoplastic cells. More sensitive methods such as RT-PCR are best used in cases with <1% FISH-positive cells, and competitive RT-PCR methods are best used when standard RT-PCR assays are negative. Competitive RT-PCR is the most sensitive technique for detection of minimal residual disease *(72)*.

3.7. IMATINIB MESYLATE The discovery of *bcr-abl* as an oncoprotein with tyrosine kinase activity has led to the development of the first Food and Drug Administration (FDA)-approved molecularly targeted therapeutic agent for a hematologic neoplasm, that being CML. Imatinib mesylate (Gleevec™, STI571; Novartis Pharmaceutical Corp., East Hanover, NJ) is a 2-phenylaminophyrimidin that acts as a competitive inhibitor at the ATP-binding site of *abl* within *bcr-abl*, with subsequent inhibition of phosphorylation of tyrosine residues of *bcr-abl* and its substrates. Via this mechanism, imatinib mesylate specifically inhibits the tyrosine kinase activity of leukemic cells and spares normal cells *(73)*

Imatinib mesylate has been shown to induce durable clinical remissions in most patients with CML in chronic phase and is the most effective single drug for this patient group *(74)*. Conventional cytogenetic responses to imatinib mesylate are semiquantified as minimal, minor, major, or complete when the percentage of Ph-negative metaphases is 1–32%, 33–66%, 67–99%, or 100%, respectively. In 95% of chronic-phase CML patients, a variable degree of cytogenetic response is achieved with imatinib mesylate therapy alone.

Combined with other chemotherapy agents, imatinib mesylate is also transiently effective in some patients with accelerated- or blast-phase CML, as well as patients with precursor B-lymphoblastic leukemia carrying the t(9;22).

3.8. RESISTANCE TO IMATINIB MESYLATE Relapse occurs in approx 15% of patients with chronic-phase CML and nearly 100% of patients with accelerated- or blast-phase CML during treatment with imatinib mesylate *(75)*. Multiple mechanisms of resistance have been hypothesized on the basis of in vitro and animal experiments, including clonal evolution, *bcr-abl* amplification, *bcr-abl* mutations, binding of imatinib mesylate to the plasma protein α1 acid glycoprotein, and development of a multidrug resistance phenotype. Of these, in vivo point mutations in *bcr-abl* appear to be a common mechanism of resistance *(76,77)*. These mutations result in amino acid substitutions at the tyrosine kinase domain of *abl*, decreasing the affinity of leukemic cells for binding with imatinib mesylate *(76)*. Up to 18 candidate mutations in *bcr-abl* have been reported to date, with the most common point mutation being isoleucine substituted for threonine at position 315 (T315I). These mutations have been detected in patients with precursor-B-lymphoblastic leukemia/lymphoma as well as those with CML. Whether these mutations develop during therapy or are present in a pre-existing subclone of leukemic cells at time of initial diagnosis that is selected after treatment is not entirely clear at this point in time. However, an analogy to antibiotic resistance seems likely and, therefore, suggests that *bcr-abl* mutations occur spontaneously and are selected by therapy with imatinib mesylate.

Elevated *bcr-abl* expression secondary to gene amplification of *bcr-abl* also has been shown in a small subset of CML cases *(76)*. Gene amplification is thought to provide additional survival and proliferative advantage to the leukemic cells.

4. CHRONIC MYELOMONOCYTIC LEUKEMIA

Chronic myelomonocytic leukemia (CMML) can be a relatively difficult disease to classify because these neoplasms can have both myelodysplastic and myeloproliferative features, with one or the other feature predominating in an individual case. As a result, in some studies, these cases have been classified as atypical or Ph-negative CML or as an unclassifiable myeloproliferative disorder. The WHO Classification recognized the difficulty in classifying these neoplasms by placing CMML in the category of myelodysplastic/myeloproliferative disorders *(3)*.

One subgroup of CMML cases is of particular interest for its molecular findings. A subset of CMML cases carry translocations that involve the platelet-derived growth factor beta (*pdgfb*) receptor gene at chromosome 5q33. Patients with abnormalities involving the *pdgfb* gene often have marked eosinophilia in addition to monocytosis and, thus, some cases have been classified in the past as chronic eosinophilic leukemia or idiopathic hypereosinophilic syndrome.

The most common and well-characterized translocation in patients with *pdgfb* gene abnormalities is the t(5;12)(q33;p13). In this translocation, the *etv6* (*ets*-type variant 6) gene, also known as *tel* (translocated ets leukemia) at chromosome 12p13 is juxtaposed with the *pdgfb* gene, resulting in a *etv6-pdgfb*

fusion gene on the derivative chromosome 5 that constitutively activates the tyrosine kinase activity of *pdgfb (78,79)*. This translocation contains the helix–loop–helix domain of *etv6* and the entire tyrosine kinase and transmembrane domains of *pdgfb*.

The true frequency of the t(5;12) in CMML is unknown, as routine testing for this translocation is not performed in most molecular diagnostic laboratories. However, the frequency appears to be low, estimated to be 1–2% of patients with CMML in the WHO Classification *(3)*. Steer and Cross *(80)* reviewed the literature in 2002 and they identified only 34 cases with the t(5;12), as well as a smaller number of cases with translocations involving *pdgfb*, but without involvement of *etv6*. However, as shown by others, cases of CMML can have *pdgfb* gene rearrangements with a normal karyotype, and, thus, assessment by conventional cytogenetics underestimates the true frequency.

The t(5;12) and other translocations involving *pdgfb* has recently received much clinical interest because cells expressing ETV6-PDGFB protein respond to tyrosine kinase inhibitors in vitro *(81)*. Furthermore, patients with CMML shown to carry the t(5;12) have been shown to respond to imatinib mesylate, with both clinical and cytogenetic remission *(79)*.

4.1. VARIANT *PDFGB* TRANSLOCATIONS ASSOCIATED WITH CHRONIC MYELOMONOCYTIC LEUKEMIA

4.1.1. t(5;17)(q33;q13) The t(5;17) has been reported in rare cases of CMML. This is a reciprocal translocation involving *pdgfb* and the *rabaptin-5* gene at chromosome 17q13 *(82)*. Patients with CMML associated with the t(5;17) have been shown to respond to imatinib mesylate.

4.1.2. t(5;7)(q33;q11.2) The t(5;7) fuses the Huntingtin interacting protein 1 (*hip1*) gene at 7q11.2 with the *pdgfb* gene *(83)*. Treatment with imatinib mesylate in patients with CMML that carry this translocation has not been reported to date.

4.1.3. t(5;10)(q33;q22) The t(5;10) fuses the H4(D10S170) gene at 10q22 with the *pdgfb* gene *(84)*. Treatment with imatinib mesylate in patients with CMML that carry this translocation has not been reported to date.

4.2. OTHER TRANSLOCATIONS Other translocations that involve the *pdgfb* gene have been reported in the literature, mostly as single case reports. The t(5;14)(q33;q32) in which the *cev14* gene at 14q33 was fused with the *pdgfb* gene was reported in a case of AML *(85)*. However, this patient also had hepatosplenomegaly and eosinophilia, suggesting that it may have arisen from CMML. Other translocations involving the fusion of *pdgfb* with unknown partner genes are likely, as some patients reported in the literature who had CMML without cytogenetic abnormalities have responded to imatinib mesylate. Furthermore, using FISH disruption of the *pdgfb* gene has been identified in CMML cases with translocations involving 5q33 and other loci, such as 1q21, 1q22, 3p21, and 14q24 *(86)*.

4.3. RAS MUTATIONS *Ras* gene mutations, as described earlier in this chapter in the section on AML, are relatively common in CMML, reported in up to approximately half of cases in one study *(87)*.

5. CHRONIC EOSINOPHILIC LEUKEMIA

Chronic eosinophilic leukemia, and its distinction from hypereosinophilic syndrome (HES), is poorly defined in the literature. In the WHO Classification *(3),* idiopathic HES is defined as persistent eosinophilia greater than 1.5×10^9/L associated with organ involvement or dysfunction without a known cause. Obviously, all known reactive and neoplastic causes of eosinophilia are excluded from the category of idiopathic HES.

Cases of chronic eosinophilic leukemia are excluded from the category of idiopathic HES because they have evidence pointing to a neoplastic process. This evidence can be morphologic (elevated blast count) or at the molecular level, the latter shown by the results of conventional cytogenetics, FISH, or molecular techniques *(3)*. As new discoveries or the development of more sophisticated molecular techniques allow for the detection of monoclonality in cases thought to be idiopathic HES in past decades, the unfortunate result is that cases previously classified as idiopathic HES in the literature are a heterogeneous group and include cases that today would be classified more specifically.

An example of this reclassification is the recent detection of a fusion between the fip1-like-1 gene (*fip1l1*) and platelet-derived growth factor receptor alpha (*pdfgα)* in patients with HES *(88)*. With the detection of this fusion gene, these cases are known to be monoclonal and would be reclassified as chronic eosinophilic leukemia according to the criteria of the WHO Classification. The discovery of this fusion gene is also of great clinical interest because affected patients respond to imatinib mesylate

Both the *fip1-like-1* and *pdgfα* genes are located at chromosome 4q12 and are approx 800 kb apart. Most commonly, *fip1-like-1* and *pdgrα* are fused as a result of an interstitial deletion (Fig. 13) and these neoplasms have a normal karyotype. The rare t(1;4)(q44;q12) also has been reported that generates the *fip1-like-1-pdgfα,* also as a result of a chromosome 4 interstitial deletion. The *fip1-like-1* gene has a sequence similar to the *fip1* gene in *Saccharomyces cerevisiae* and encodes encodes a protein essential to the process of polyadenylation in that species. The *pdgfα* gene encodes a class III tyrosine kinase. The breakpoints in *fip1-like-1* are relatively scattered thoughout exons 7–10. The breakpoints in *pdgfα* are relatively clustered within exon 12.

The true frequency of the *fip1-like-1-pdgfα* fusion gene is unknown, but the frequency might be as high as 50% of HES cases. In a study of 16 cases of HES by Cools and colleagues, nine cases (56%) carried the *fip1-like-1-pdgfα* fusion gene *(88)*. In that study, nested RT-PCR was used to detect the fusion gene.

6. ACUTE LYMPHOBLASTIC LEUKEMIA

Acute lymphoblastic leukemia (ALL) and lymphoblastic lymphoma (LBL) are morphologically and immunophenotypically similar neoplasms, although ALL patients present predominantly with peripheral blood and bone marrow involvement, and LBL patients present initially with prominent lymphadenopathy and a mediastinal mass, with absent or low-level leukemic involvement *(3)*.

Cases of ALL/LBL are neoplasms of immature (precursor) lymphoid cells. Their lineage is known to be immature by the expression of terminal deoxynucleotidyl transferase (TdT) and CD34 (frequent) and the absence of mature lymphoid antigens. These cases can be further divided into precursor (pre) B-cell and T-cell types. At the immunophenotypic and molecular level, cases of pre-B ALL and LBL are very similar, as are cases of pre-T-ALL and -LBL. However, the immunophenotypic and

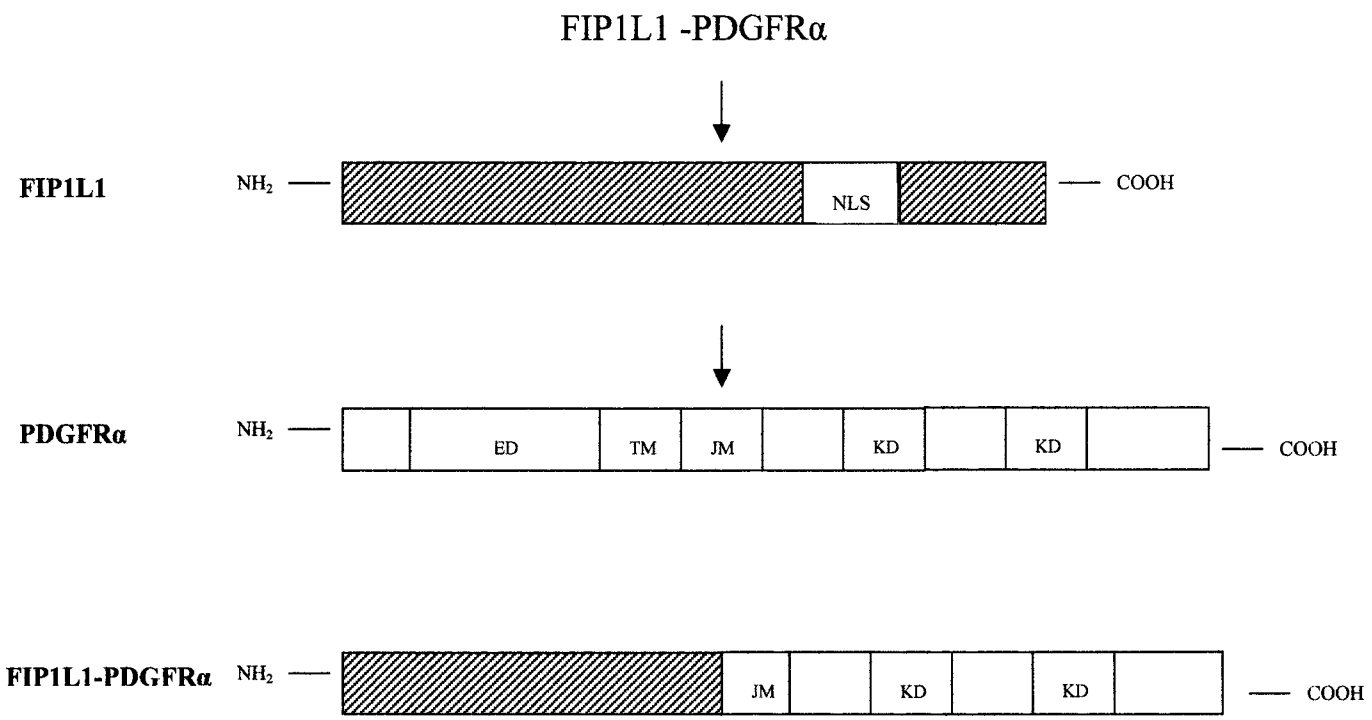


Fig. 13. Schematic of FIP1-Like-1 and PDGFα proteins and the FIP1L1-PDGFα fusion protein that is common in idiopathic hypereosinophilic syndrome (see text). The fusion is usually the result of an interstitial deletion on chromosome 4. NLS, nuclear localization signal; ED, extracellular domain; TM, transmembrane domain; JM, juxtamembrane domain; KD, kinase domain. Arrows indicate breakpoints.

molecular findings in pre-B and pre-T neoplasm are distinctive. Thus, in the WHO Classification, these tumors are designated separately as precursor B-cell or T-cell lymphoblastic leukemia/lymphoblastic lymphoma *(3)*.

Patient age adds further to the heterogeneity of ALL/LBL. For example, pre-B-ALL/LBL is primarily a disease of children and usually presents as leukemia (rather than lymphoma) *(3)*. However, a second, lower peak in ALL incidence occurs in adults approx 50 yr of age. There are number of molecular abnormalities described in pre-B-ALL, and the distribution of the frequency of these abnormalities is unequal in pediatric and adult patients. Most likely, these differences in distribution account for the disparate clinical outcomes between children and adults with pre-B-ALL.

Similarly, patients with pre-T-ALL/LBL are most often adolescents or young adults who present with lymphoma (i.e., medistinal mass and lymphadenopathy). However, a second peak in disease incidence occurs in older adults as ALL, and the distribution of molecular abnormalities (although relatively less well studied) is different between these two age groups.

A large number of cytogenetic and molecular abnormalities have been described in pre-B- and pre-T-ALL/LBL (Fig. 14) *(89)*. However, a subset of ALL/LBL cases currently do not have known distinctive molecular abnormalities. Thus, as knowledge expands in ALL/LBL, many other molecular abnormalities are likely to be discovered. In addition, recent gene expression profiling studies have shown similar patterns of gene expression in ALL/LBL cases with and without known distinctive chromosomal translocations *(5,90)*. These results suggest that many cases of ALL/LBL without known molecular abnormalities have lesions that activate the same pathways as in ALL/LBL cases with known molecular abnormalities. Here, we discuss some of the most common abnormalities, known in pre-B- and pre-T-ALL/LBL.

6.1. PRECURSOR B-CELL

6.1.1. DNA Chromosome Number (Ploidy) DNA chromosome number has been shown in numerous studies to be of prognostic significance *(89)*. High hyperdiploidy, defined as 51–65 chromosomes, occurs in approx 30% of childhood and 5–10% of adult pre-B-ALL and is associated with a better patient prognosis. The chromosomes most often duplicated are 4, 6, 10, 14, 17, 18, 21, and X. DNA hypodiploidy, defined as 45 chromosomes or less, is less common in ALL. A chromosome number of 33–44 is rare, occurring in less than 1% of pediatric and adult ALL cases, but these patients have a poorer prognosis, and rare patients with 23–29 chromosomes (so-called near-haploidy) have an extremely poor prognosis.

6.1.2. t(9;22)(q34;q11) The t(9;22)(q34;q11.2) has been discussed in detail in the section on CML. The t(9;22) occurs in approx 20% adult and 3–4% of pediatric pre-B-ALL cases *(89,91)*. In adults, the frequency of the t(9;22) in pre-B-ALL correlates with increasing patient age.

Patients with pre-B-ALL associated with t(9;22)(q34;q11) are characterized by marked leukocytosis, high blast counts, and a tendency for central nervous system involvement at time of diagnosis. Historically, both adult and pediatric patients with t(9;22)-positive pre-B-ALL have had a poorer prognosis than patients with t(9;22)-negative ALL. However, more aggressive chemotherapy regimens and stem cell transplantation might be improving the prognosis of patients with t(9;22)-positive ALL *(92)*.

In both adult and childhood cases of ALL, the breakpoints in the *abl* gene usually occur between exons 1 and 2, the same region

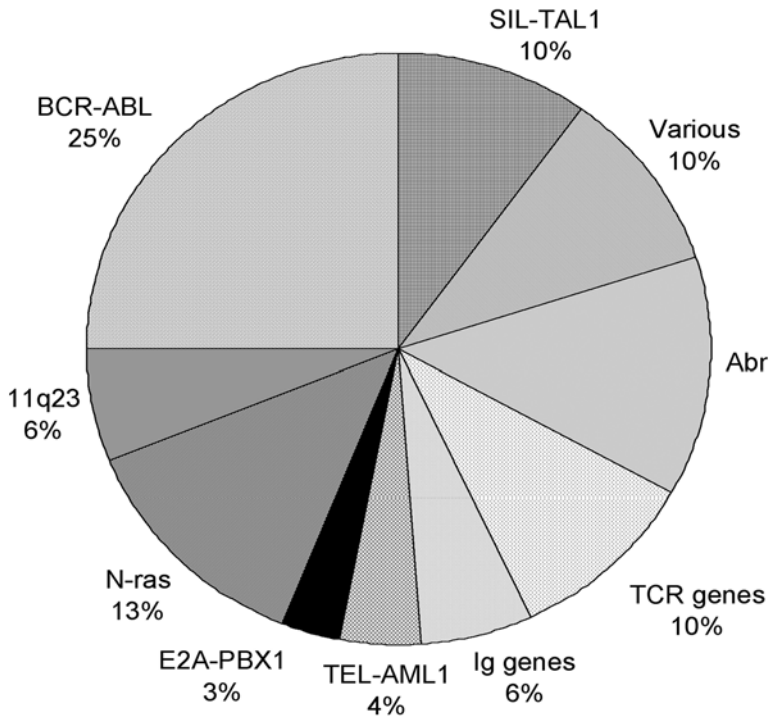


Fig. 14. Estimated prevalence of molecular lesions in adult ALL.

as in CML *(91)*. By contrast, the breakpoints in the *bcr* gene commonly involve the m-*bcr* region (between exons 1 and 2) and result in a *bcr-abl* fusion gene that encodes a 7.5-kb mRNA and a p190 protein. Approximately 50% of adults and 80% of childhood pre-B-ALL cases encode for p190. The remaining cases of pre-B-ALL have breakpoints in the M-*bcr* region of the *bcr* gene, creating a *bcr-abl* fusion gene than encodes a p210 protein similar to that in CML. The p190 protein has stronger tyrosine kinase activity than p210 protein *(91)*. Unlike CML, deletions of chromosome 9 are uncommon in pre-B-ALL cases *(89)*.

6.1.2.1. Molecular Methods of Detection Both the p190 and p210 variants of the t(9;22) can be detected by FISH and RT-PCR based assays. Southern blot analysis, although much more inconvenient, also can be used.

6.1.3. t(12;21)(p12;q22) The t(12;21) is the most common recurrent translocation in pediatric pre-B-ALL, occurring in approx 25% of cases *(3,89,93)*. This translocation is uncommon in adult pre-B-ALL, in less than 5% (or less) of cases. Pediatric patients with t(12;21)-positive pre-B-ALL have a favorable prognosis.

The t(12;21) fuses the N-terminus of the *etv6* gene (also known as *tel*) on chromosome 12, with the *runx1* gene (also known as *aml1*) on chromosome 21, generating a *etv6-runx1* fusion gene (Fig. 15) *(93)*. The reciprocal *runx1-etv6* fusion gene is also created, but only in a subset of cases and is not involved in leukemogenesis.

The breakpoints usually occur within intron 5 of *etv6* and intron 1 of *runx1*. A common splice variant involves exon 5 of *etv6* and exon 3 of *runx1*. The precise mechanism by which *etv6-runx1* induces leukemia is unclear, but molecular studies suggest that the translocation occurs prenatally at the hematopoietic stem cell level and that additional postnatal mutations/abnormalities are required for leukemic transformation *(94)*. The unrearranged etv6 allele in t(12;21)-positive pre-B-ALL is often deleted, suggesting that *etv6* functions as a tumor suppressor gene *(94)*.

The normal product of the *etv6* gene is a nuclear phosphoprotein with a helix–loop–helix domain that is a member of the *ets* family of transcription factors (Fig. 16) *(93)*. The *runx1* gene encodes a DNA-binding subunit of the AML-1/CBFβ transcription factor complex.

6.1.3.1. Molecular Methods of Detection The t(12;21)(p12;q22) is difficult to recognize by conventional karyotype analysis, perhaps because affected loci on chromosomes 12 and 21 are morphologically similar. The t(12;21) is detectable by FISH and PCR methods.

6.1.3.2. Variant Translocations Involving etv6 in Pre-B-ALL Translocations involving *etv6* have been reported in rare cases of pre-B-ALL, including the t(9;12)(q11;q13) involving the *pax-5* gene at chromosome 9p11 and the t(9;12)(p24;p13) involving the *jak-2* gene at chromosome 9p24 *(93)*.

6.1.4. t(1;19(q23;p13) The t(1;19)(q23;p13) is present in approx 6% of pediatric pre-B-ALL, but is less common in adults *(89,95)*. The t(1;19) is present in approx 25% of pre-B-ALL cases that are positive for cytoplasmic IgM. Historically, the t(1;19) in pre-B-ALL has been associated with a poor prognosis, but more aggressive chemotherapy in more recent studies has resulted in improved prognosis.

The t(1;19) splits the *e2a* gene on chromosome 19q13 and the *pbx1* gene on chromosome 1q23, resulting in the *e2a-pbx1* fusion gene on the derivative chromosome 19 *(95)*. The *e2a* gene encodes two transcription factors, E12 and E47, necessary for normal hematopoiesis and regulation of B-cell development. The *pbx1* gene is a homeobox gene.

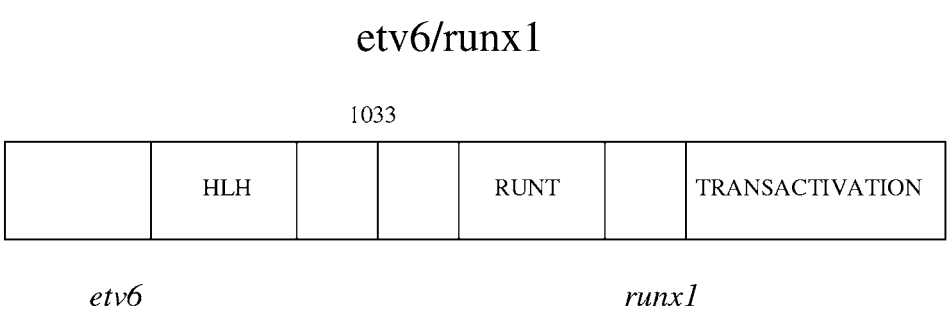


Fig. 15. Schematic of the *etv6-runx1* fusion gene created as a result of the t(12;21)(p12;q22) that is common in childhood pre-B-ALL. The breakpoint in *etv6* occurs at nt 1033 and fuses the first 5 exons of *etv6* in frame with exon 2 of *runx1*, HLH, helix–loop–helix; RUNT, runt homology.

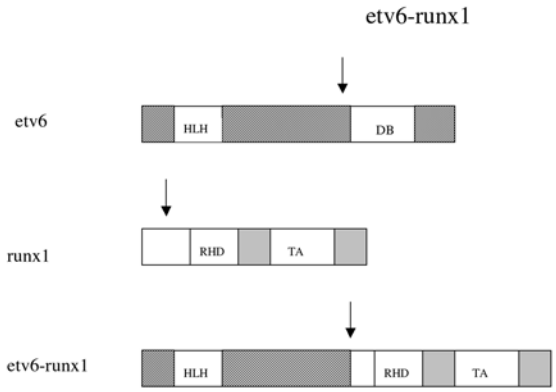


Fig. 16. Schematic of the normal ETV6 and RUNX1 proteins and the chimeric ETV6-RUNX1 encoded as a result of the t(12;21)(p12;q22). HLH, helix–loop–helix domain; DB, DNA-binding domain; RHD, runt homology domain; TA, transactivation domain.

6.1.4.1. Molecular Methods of Detection The t(1;19) can be detected by conventional cytogenetics. The translocation also can be detected by FISH and PCR methods because the breakpoints are clustered.

6.1.4.2. Variant Abnormalities Involving the e2a Gene The t(17;19)(q22;p13) is a rare translocation in which the *e2a* gene is disrupted and fused with the *hlf* gene on chromosome 17. A cryptic inversion of chromosome 19 involving *e2a*, the inv(19)(p13q13) that results in an *e2a-fb1* fusion gene, also has been described *(96)*.

6.1.5. Gene Mutations Relatively less is known about gene mutations in pre-B-ALL. The *atm* gene (ataxia telangiectasia mutated) at chromosomal 11q22-23 can be lost or mutated in approx 20% of cases *(98)*. The evidence for this is based on conventional cytogenetic and PCR studies, the former showing deletions of 11q and the latter using microsatellite markers to show loss of loci at 11q22-23. *ras* gene mutations, most often in codons 12 and 13 of *n-ras*, have been reported in approx 10–15% of pre-B-ALL cases *(98,99)*. The presence of these mutations has been correlated with poorer prognosis. *p16* gene mutations are reported in a small subset of pre-B-ALL cases, but are much more common pre-T-ALL cases.

6.2. PRECURSOR-T CELL

6.2.1. Tal-1 The most common genetic abnormality in pre-T-ALL, occurring in approx 30% of cases, involves the t*al-1* gene on chromosome 1p32 *(100)*. The *tal-1* gene is usually structurally altered, via interstitial deletion (25%) or translocation (<5%), but biallelic activation of *tal-1* can also occur *(101,102)*. Patients with pre-T-ALL and *tal-1* abnormalities tend to have prominent extramedullary disease *(100)*.

In most affected pre-T-ALL cases, the *tal-1* gene is rearranged via an interstitial deletion of chromosome 1p32 *(100,102)*. This results in *tal-1* being joined with the *sil* gene by illegitimate VDJ recombination. Much less often, the *tal-1* gene is juxtaposed with the enhancer of the T-cell-receptor δ- or β-chain genes, via the t(1;14)(p32;q11) or t(1;7)(p33;q35), respectively; *tal-1* expression is then dysregulated. In mouse models, *tal-1* overexpression has been shown to activate NF-κB, causing developmental arrest *(103)*.

6.2.2. t(5;14)(q35;q32) The t(5;14)(q35;q32) is present in approx 25% of pre-T-ALL cases and is more frequent in children than adults (104,105). The breakpoints on chromosome 5 are relatively clustered and disrupt a gene, *ranbp17*, but this gene is not thought to be involved in pathogenesis. Instead, a nearby gene is upregulated, *hox11-like-2*. The breakpoints on chromosome 14 are more widely scattered, with the closest gene being *ctip2/bcl-11b* that is known to be highly expressed during T-cell differentiation *(104,105)*.

The t(5;14) was detected using a FISH approach, and cases may be missed by conventional cytogenetics. H*ox11-like-2* gene overexpression appears to be restricted to cases with the t(5;14) and can be used as a surrogate for the translocation *(105)*.

Rare variant translocations involving the *hox11-like 2* gene, the t(5;7)(q35;q21), and the t(2;5)(p21;q35), also have been described *(105)*.

6.2.3. Hox11 The *hox11* gene, a homeobox gene involved in transcriptional regulation, is activated in approx 5–10% of pre-T-ALL. Patients with pre-T-ALL associated with *hox11* gene activation have a favorable prognosis *(106)*

The *hox11* gene is most often activated by translocation, either the t(10;14)(q24;q11) or t(7;10)(q34;q24), in which *hox11* is juxtaposed with the enhancer of T-cell-receptor α/δ locus at 14q11 or the T-cell-receptor β-chain gene at 7q34.

6.2.4. Gene Mutations Deletions of chromosome 9p21–22 have been shown by conventional cytogenetics in approx 10% of cases of pre-T-ALL *(107)*. These patients have been associated with high leukocyte and blasts counts and a poorer prognosis.

The 9p21-22 locus is the site of the $p16^{INK4a}$ (*mts1*), $p15^{INK4b}$ (*mts2*), and $p14^{ARF}$ genes and mutations in these genes are common in pre-T-ALL, in up to 80% of pediatric pre-T-ALL cases *(108)*. These genes encode proteins that are inhibitors of

cyclin D-cyclin dependent kinase 4/6 complexes, and thus mutations lead to disrupted cell cycle regulation.

6.3. *mll* GENE ABNORMALITIES The *mll* gene (mixed lineage leukemia) has been discussed previously in the section on AML. The *mll* gene is also involved in a subset of cases of ALL of either pre-B or pre-T lineage *(89)*. Both *mll* gene duplications and translocations have been described.

The *mll* gene translocations are very common in leukemias in infants less than 1 yr of age. These leukemias can have an immunophenotype that is myeloid (AML), lymphoid (ALL), or biphenotypic, and affected infants often have a poor prognosis regardless of the immunophenotype. There are up to 50 translocations involving the *mll* gene described in the literature and the partner genes are currently known in a subset of cases *(89)*. The t(4;11)(q21;q23) is most common in ALL *(109)*.

6.3.1. t(4;11)(q21;q23) The t(4;11) is the most common translocation involving the *mll* gene identified in cases of ALL, found in approx 2–3% of all pediatric cases and more commonly in neonatal cases *(109)*. The t(4;11) also has been detected in approx 5% of adult pre-B-ALL. Immunophenotypically, the blasts of t(4;11)-positive ALL often lack CD10 and are positive for the myelomonocytic marker CD15. These neoplasms also have the potential to transform to acute monoblastic leukemia after chemotherapy with pre-B-ALL regimens, supporting the concept that this type of acute leukemia arises from a primitive progenitor with potential of both lymphoid and monocytic differentiation. These neoplasms are classified as acute leukemia of ambiguous lineage in the WHO Classification *(3)*.

The t(4;11) fuses the 5′ portion of the *mll* gene with the 3′ portion of the *af4* gene (also known as *fel*) on chromosome 4q21. The *mll-af4* fusion gene resides on the derivative chromosome 11. Many other translocations that involve the *mll* gene are also reported in pre-B-ALL cases, such as the t(1;11)(p32;q23) involving the *af-1p* gene, t(2;11)(q11;q23) involving the *af4-related* gene, t(6;11)(q27;q23) involving the *af6* gene, the t(9;11)(p22;q23) involving the *af9* gene, the t(10;11)(p12;q23) involving the *af10* gene, and the t(11;19)(q23;p13.3) the *enl* gene *(89,110,111)*. Each of these translocations is present in 1% or less of pre-B-ALL cases.

6.3.2. Pre-T-Cell ALL with mll Translocations Although least common, *mll* gene translocations also occur in cases of pre-T-ALL *(89,112)*. Although these translocations also result in upregulation of *hox* genes, similar to pre-B-ALL cases, cases of pre-T-ALL uncommonly express myeloid antigens, unlike their pre-B-cell counterparts.

7. CHRONIC LEUKEMIAS OF LYMPHOID ORIGIN

Although chronic lymphocytic leukemia is, by far, the most common type of chronic leukemia of lymphoid origin, there are other types of chronic B-cell or T-cell leukemia. However, many of these neoplasms also involve tissues and are, therefore, considered as lymphomas with a prominent leukemic component (e.g., mantle cell lymphoma, Burkitt lymphoma, follicular lymphoma, etc.). As lymphomas in leukemic phase have been discussed in detail in the chapter on lymphomas, they are not discussed here.

7.1. CHRONIC LYMPHOCYTIC LEUKEMIA Chronic lymphocytic leukemia (CLL) and small lymphocytic lymphoma (SLL) are morphologically and immunophenotypically similar neoplasms, although CLL patients present predominantly with peripheral blood and bone marrow involvement and SLL patients present with lymphadenopathy. CLL and SLL are lumped together in the WHO Classification *(3)*. CLL is the most common type of leukemia in adults and SLL represents approx 6% of all non-Hodgkin's lymphomas in the United States and Europe.

7.1.1. Conventional Cytogenetics and FISH Results Using conventional cytogenetic analysis, chromosomal abnormalities have been identified in approx 50% of cases of CLL/SLL *(113)*. Poor growth of the neoplastic cells is the likely explanation for the high frequency of normal karyotypes. However, using FISH methods, approx 70–80% of CLL/SLL cases have abnormalities *(114,115)*.

Deletions of 13q14, either one or both alleles, are the most common abnormalities detected in CLL/SLL, identified in approx 50% of cases by FISH *(114,115)*. The gene involved at this locus, hypothesized to be a tumor suppressor gene, is unknown. The *rb* gene at 13q14 is not involved, as the minimally deleted region is 1.6 CM telomeric to the *rb* gene. The presence of 13q14 deletions correlates with a favorable clinical course.

Chromosome 11q deletions, most often involving 11q22-23, occur in 10–20% of patients with CLL/SLL cases *(114)*. The ataxia–telangiectasia mutated (*atm*) gene located at this locus is mutated in a subset of these patients. However, other genes at this locus could also be involved. 11q deletions are associated with younger patient age, more advanced clinical stage, extensive lymph node involvement, and a poorer clinical course.

Trisomy 12 is found in approx 10–20% of CLL/SLL cases *(114)*. The minimally duplicated segment of chromosome 12 is 12q13-q21.2. The gene (or genes) presumably amplified and hypothesized to be an oncogene as a result of trisomy 12 is unknown. Trisomy 12 has been correlated with atypical morphologic and immunophenotypic features in CLL patients and with poorer prognosis in some studies.

Chromosome 17p13 deletions occur in 10–15% of CLL/SLL cases and these deletions usually involve the *p53* gene *(114)*. Chromosome 17q deletions or *p53* gene mutations strongly correlate resistance to therapy and shorter survival. Deletions of chromosome 6q occur in approx 5% of CLL/SLL cases and are associated with higher leukocyte counts and more extensive lymphadenopathy, but have not been shown to correlate with prognosis *(114)*.

7.1.2. Chromosomal Translocations Chromosomal translocations have been identified in a small subset of CLL/SLL. In less than 5% of CLL/SLL cases, the *bcl*-2 gene at 18q21 is rearranged as a part of the t(2;18)(p11;q21) or t(18;22)(q21;q11), involving the *Igκ* (chromosome 2p11) or *Igλ* (chromosome 22q11) gene loci *(116)*.

The t(14;19)(q32,q13) is a rare cytogenetic abnormality, found in less than 1% of cases of CLL/SLL, and is associated with young patient age at time of initial diagnosis, atypical cytologic features, and poor prognosis *(117)*. The t(14;19) juxtaposes the *bcl*-3 gene at 19q13 with *IgH* in a head-to-head configuration, resulting in overexpression of *bcl-3* because of the 3′ *IgH* enhancer *(118)*. The *bcl-3* gene encodes a member of the IκB family of proteins that are known to inhibit NFκB. However, unlike all other IκB proteins, *bcl-3* is a coactivator of transcription.

The t(2;14)(p13;q32) is a rare, recently described translocation that has been described in cases of adult CLL/SLL, lymphoplasmacytic lymphoma, and so-called childhood CLL. The t(2;14) juxtaposes the *bcl-11A* gene at 2p13 with *IgH* at 14q32. The *bcl-11A* gene is the human homolog of the mouse *evi-9* gene, which is deregulated in murine myeloid leukemias after proviral integration *(119)*. In one report, four neoplasms with the t(2;14) lacked somatic mutations of the *IgH* variable region genes *(120)*.

The t(11;14)(q13;q32) has been reported in cases of CLL/SLL in the literature. However, in retrospect, it seems likely that most of these cases were examples of mantle cell lymphoma in leukemia phase *(121)*.

7.1.3. Somatic Hypermutation of the Immunoglobulin Genes Somatic hypermutation of the IgH and Ig light-chain variable region genes is a physiologically normal process in which single nucleotides are exchanged, or small segments of DNA are inserted or deleted. Somatic mutation takes place primarily in the germinal center after exposure to antigen. Somatic mutation is advantageous to the organism because, over time, the Ig genes encode antibodies that bind with antigen more strongly are selected. This mechanism explains the greater antibody affinity known to occur in the anamnestic immune response.

Somatic mutation is known to occur in the rearranged IgH and Ig light-chain V genes of a subset of CLL/SLL cases. Using the arbitrary cutoff of greater than or equal to 2% mutations after analysis by sequencing, approx 60% of CLL/SLL cases carry mutated IgH V genes *(122,123)*. Mutation status has been shown to independently predict patient survival, as CLL/SLL patients with unmutated IgH V genes have a significantly worse prognosis *(124)*.

7.1.4. Gene Expression Profiling Recent studies using gene expression profiling methods have suggested that all CLL/SLL cases share a common gene expression profile related to memory B-cells *(125,126)*. Compared with other types of B-cell non-Hodgkin's lymphoma (e.g., follicular lymphoma), CLL/SLL cases exhibit downregulation of most genes involved in cell proliferation, including those involved in the cell cycle, DNA synthesis, and DNA replication. Genes involved in inhibiting apoptosis are also upregulated.

7.1.5. Transformation Approximately 5–10% of patients with CLL/SLL develop transformation to a higher-grade lymphoid neoplasm involving tissue specimens, known by the eponym Richter's syndrome *(127)*. Most often, CLL/SLL patients develop diffuse large B-cell lymphoma, but less commonly, patients can develop classical Hodgkin's disease. In the peripheral blood, patients with CLL can evolve to a neoplasm characterized by numerous prolymphocytes, known as prolymphocytoid transformation of CLL.

It is estimated that the diffuse large B-cell lymphoma is clonally related to CLL/SLL in approx 60% of cases *(127)*. The genetic events that result in Richter's syndrome are poorly understood. Mutations of the *p16* and *p53* genes have been identified in a subset of diffuse large B-cell lymphoma specimens arising in CLL/SLL patients. Bea and colleagues *(128)* used comparative genomic hybridization to study both the CLL and diffuse large B-cell lymphoma components of Richter's syndrome cases and showed gains of 2pter and 7pter and losses of 8p, 11q, and 17p in the diffuse large B-cell lymphomas.

In contrast, cases of CLL in prolymphocytoid transformation are usually clonally related to the prior CLL. Cytogenetic abnormalities of chromosome 17p and *p53* gene mutations have been identified in a substantial subset of cases *(129,130)*. Translocations involving *c-myc* also have been reported in a small number of cases *(131)*.

7.2. B-CELL PROLYMPHOCYTIC LEUKEMIA B-cell prolymphocytic leukemia (PLL), as the entity is described classically, is characterized by a high peripheral blood leukocyte count, prominent splenomegaly with minimal lymphadenopathy, and an aggressive clinical course *(132)*. In the peripheral blood and bone marrow, numerous prolymphocytes are present. Prolymphocytes are of medium cell size with central nuclei, prominent central nucleoli with perinucleolar chromatin condensation, and abundant pale cytoplasm.

In recent years, the entity of B-cell PLL has come under scrutiny. Many cases previously classified as B-cell PLL were likely to be cases of CLL in prolymphocytoid transformation. In the WHO Classification, such cases are now clearly excluded from the category of B-cell PLL *(3)*. In earlier studies using conventional cytogenetic methods, up to 30% of cases were shown to carry the t(11;14). These t(11;14)-positive cases were probably variants of mantle cell lymphoma in the leukemic phase *(3,129)*.

Molecular studies of B-cell PLL have shown a high frequency of *p53* gene mutations, approx 50% *(130,132)*. In addition, conventional cytogenetic analysis has revealed complex karyotypes in many cases.

7.3. T-CELL PROLYMPHOCYTIC LEUKEMIA T-cell prolymphocytic leukemia (T-PLL) is an uncommon type of leukemia, although it is the most common leukemia of mature T-cell lineage. Clinically, patients most often present with a rapidly rising leukocyte count and hepatosplenomegaly, but a subset of patients initially have an indolent course *(133)*. Morphologically, the leukemic cells usually resemble prolymphocytes, although a small cell variant with a small or indistinct nucleolus also has been described. Cases of T-cell chronic lymphocytic leukemia reported in the literature also may be examples of the small cell variant of T-PLL *(133)*.

7.3.1. Rearrangements of 14q32.1 The most common chromosomal abnormality in T-PLL is rearrangement of 14q32.1, including inv(14)(q11;q32) and the rare t(14;14)(q11;q32) and t(7;14)(q35;q32) *(134)*. These abnormalities juxtapose the T-cell leukemia/lymphoma-1 (*tcl-1*) locus at 14q32 adjacent to the T-cell-receptor (TCR) α/δ locus at 14q11 in the inv(14) and t(14;14), or adjacent to the TCRβ gene in the t(7;14). The breakpoints involving 14q32.1 occur over a 120-kb stretch of DNA. Another rare translocation that occurs in T-PLL is the t(X;14)(q28;q11), in which the *mtcp1* gene at Xq28 is juxtaposed with the TCR γ/δ locus. In these abnormalities the TCRα enhancer, or TCRβ in the rare t(7;14), upregulate expression of *tcl-1* or *mtcp-1*. TCL-1 and MTCP-1 proteins are 41% homologous and thus MTCP-1 is considered a member of the TCL-1 family *(134)*.

The *tcl-1* locus at 14q32.1 includes at least four genes: *tcl-1*, *tcl-1b*, and the *tcl-1* neighboring genes (*tng-1* and *tng-2*). *tcl-1b* is 60% homologous with *tcl-1*. Much less is known about *tng-1* and *tng-2;* these genes are not considered part of the *tcl-1*

family. Deregulation of *tcl-1* or *mtcp-1* in transgenic mice results in T-cell leukemia, supporting the concept that both genes are oncogenes. The normal functions of the *tcl-1* family of oncoproteins are unknown.

Inversions and translocations involving 14q32.1 result in increased levels of TCL-1 protein providing the cell with a growth advantage. The protein is small, with four exons and an open reading frame of 114 amino acids. TCL-1 is expressed normally by precursor B-cells and T-cells (CD4–CD8–), by mantle zone and germinal center B-cells, and by normal ovary tissue. TCL-1 is expressed in most cases of T-PLL, even in cases without cytogenetic evidence of 14q32.1 rearrangement *(134)*. TCL-1 has been shown to bind with and enhance the kinase activity of AKT and, thus, may act in cells via the PI3-kinase-dependent AKT/protein kinase B pathway.

7.3.2. Atm Gene Mutations Ataxi-telangiectasia (AT) is a rare autosomal recessive disease characterized by cerebellar ataxia, increased sensitivity to ionizing radiation, predisposition to leukemias (T-PLL and T-ALL) and B-cell lymphomas. The risk of leukemia or lymphoma in AT patients has been estimated to be up to 250 times the general population *(135)*.

Patients with AT have mutations in the *atm* gene located at 11q22-23. This gene encodes a nuclear phosphoprotein of approx 350 kDa that is implicated in cell cycle regulation, signal transduction, and the response to DNA damage. There is evidence to suggest that these proteins respond to DNA damage by phosphorylating one or more substrates, including P53, ABL, and replication protein A to recruit proteins to regions of DNA repair and /or to activate signal transduction pathways.

In AT patients, most (>75%) *atm* gene mutations result in a truncated ATM protein or no protein at all. Patients with AT who have homozygous abnormalities of *atm* have a markedly increased risk of T-PLL *(134,135)*. Both the inv(14)(q11q32) and t(14;14)(q11;q32) also have been observed in patients with AT, prior to onset of overt T-PLL, suggesting that other abnormalities are also involved in the pathogenesis of T-PLL in AT patients.

Mutations in the *atm* gene can be seen in 50% of cases of sporadic T-PLL. The *atm* mutations in T-PLL are primarily missense mutations that interfere with the catalytic function of *atm* or directly prevent ATP binding or substrate recognition. Usually, both alleles are deleted or mutated in T-PLL cases, suggesting the *atm* functions as a tumor suppressor gene.

REFERENCES

1. Landis, S. H., Murray, T., Bolden, S., and Wingo, P. A. Cancer statistics. *CA Cancer J. Clin.* **49:**8–31, 1999.
2. Bennett, J. M., Catovsky, D., Daniel, M.T., et al. Proposed revised criteria for the classification of acute myeloid leukemia. A report of the French–American–British Cooperative Group. *Ann. Intern. Med.* **103:**620–625, 1985.
3. Jaffe, E. S., Harris, N. L., Stein, H., and Vardiman, J. W. (eds.). *Pathology and Genetics* of *Tumours of Haematopoietic and Lymphoid Tissues. World Health Organization Classification of Tumours*. IARC, Lyon, 2001.
4. Haferlach, T., Kohlmann, A., Kern, W., et al. Gene expression profiling as a tool for the diagnosis of acute leukemias. *Semin. Hematol.* **40:**281–295, 2003.
5. Yeoh, E. J., Ross, M. E., Shurtleff, S. A., et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* **1:**133–143, 2002.
6. Neubauer, A., Dodge, R. K., George, S. L., et al. Prognostic importance of mutations in ras proto-oncogenes in *de novo* acute myeloid leukemia. *Blood* **83:**1603–1611, 1994.
7. Stirewalt, D. L., Kopecky, K. J., Meshinchi, S., et al. FLT3, RAS, and TP53 mutations in elderly patients with acute myeloid leukemia. *Blood* **97:**3589–3595, 2001.
8. Skorski, T., Kanakaraj, P., Ku, D. H., et al. Negative regulation of p120GAP GTPase promoting activity by 210Bcr/Abl: Implication of ras-dependent Philadelphia chromosome positive cell growth. *J. Exp. Med.* **179:**1855–1865, 1994.
9. Roumier, C., Fenaux, P., Lafage, M., et al. New mechanisms of AML1 gene alteration in hematological malignancies. *Leukemia* **17:**9–16, 2003.
10. Ahuja, H., Jat, P., Foti, A., et al. Abnormalities of the retinoblastoma gene in the pathogenesis of acute leukemia. *Blood* **78:**3259–3268, 1991.
11. Fenaux, P., Preudhomme, C., Quiquandon, I., et al. Mutations of the p53 gene in acute myeloid leukemia. *Br. J. Haematol.* **80:**178–183, 1992.
12. Care, R. S., Valk. P. J. M., Goodeve, A. C., et al. Incidence and prognosis of c-kit and FLT3 mutations in core binding factor (CBF) acute myeloid leukemias. *Br. J. Haematol.* **121:**775–777, 2003.
13. Stirewalt, D. L. and Radich, J. P. The role of FLT3 in haematopoietic malignancies. *Nat. Rev. Cancer* **3:**650–665, 2003.
14. Schnittger, S., Schoch, C., Dugas, M., et al. Analysis of FLT3 length mutations in 1003 patients with acute myeloid leukemia: correlation to cytogenetics, FAB subtype, and prognosis in the AMLCG study and usefulness as a marker for the detection of minimal residual disease. *Blood* **100:**59–66, 2002.
15. Thiede, C., Steudel, C., Mohr, B., et al. Analysis of FLT3-activating mutations in 979 patients with acute myelogenous leukemia: association with FAB subtypes and identification of subgroups with poor prognosis. *Blood* **99:**4326–4335, 2002.
16. Caligiuri, M. A., Schichman, S. A., Strout, M. P., et al. Molecular rearrangement of the *ALL-1* gene in acute myeloid leukemia without cytogenetic evidence of 11q23 chromosomal translocations. *Cancer Res.* **54:**370–373, 1994.
17. Schichman, S. A., Canaani, E., and Croce, C. M. Self-fusion of the ALL1 gene. A new genetic mechanism for acute leukemia. *JAMA* **273:**571–576, 1995.
18. Caligiuri, M. A., Strout, M. P., Schichman, S. A., et al. Partial tandem duplication of *ALL1* as a recurrent molecular defect in acute myeloid leukemia with trisomy 11. *Cancer Res.* **56:**1418–1425, 1996.
19. Degos, L. The history of acute promyelocytic leukemia. *Br. J. Haematol.* **122:**539–553, 2003.
20. Lin, P., Hao, S., Medeiros, L. J., et al. Expression of CD2 in acute promyelocytic leukemia correlates with short form of PML-RAR alpha transcripts and poorer prognosis. *Am. J. Clin. Pathol.* **121:**402–407, 2004.
21. Lo Coco, F., Diverio, D., Falini, B., Biondi, A., Nervi, C., and Pelicci, P. G. Genetic diagnosis and molecular monitoring in the management of acute promyelocytic leukemia. *Blood* **94:**12–22, 1999.
22. Zhong, S., Hu, P., Ye, T. Z., et al. A role for PML and the nuclear body in genomic stability. *Oncogene* **18:**7941–7947, 1999.
23. Grignani, F., de Matteis, S., Nervi, C., et al. Fusion proteins of the retinoic acid receptor-alpha recruit histone deacetylase in promyelocytic leukemia. *Nature* **391:**815–818, 1998.
24. Grimwade, D., Biondi, A., Mozziconacci, M. J., et al. Characterization of acute promyelocytic leukemia cases lacking the classic t(15;17): results of the European Working Party. Groupe Francais de Cytogenetique Hematologique, Groupe de Francais d'Hematologie Cellulaire, UK Cancer Cytogenetics Group and BIOMED 1 European Community-Concerted Action "Molecular Cytogenetic Diagnosis in Haematological Malignancies." *Blood* **96:**1297–1308, 2000.
25. Villamor, N., Costa, D., Aymerich, M., et al. Rapid diagnosis of acute promyelocytic leukemia by analyzing the immunocytochemical pattern of PML protein with the monoclonal antibody PG-M3. *Am. J. Clin. Pathol.* **114:**786–792, 2000.
26. Sainty, D., Liso, V., Cantu-Rajnoldi, A., et al. A new morphologic classification system for acute promyelocytic leukemia distinguishes

cases with underlying PLZF/RARA gene rearrangements. Group Francais de Cytogenetique Hematologique, UK Cancer Cytogenetics Group and BIOMED 1 European Community-Concerted Acion "Molecular Cytogenetic Diagnosis in Haematological Malignancies." *Blood* **96:**1287–1296, 2000.
27. Jansen, J. H., de Ridder, M. C., Geertsma, W. M., et al. Complete remisision of t(11;17) positive acute promyelocytic leukemia induced by all trans retinoic acid and granulocyte colony-stimulating factor. *Blood* **94:**39–45, 1999.
28. Khoury, H., Dalal, B. I., Nevill, T. J., et al. Acute myelogenous leukemia with t(8;21): Identification of a specific immunophenotype. *Leuk. Lymphoma.* **44:**1713–1718, 2003.
29. Swansbury, G. J., Lawler, S. D., Alimena, G., et al. Long-term survival in acute myelogenous leukemia: a second follow-up of the Fourth International Workshop on Chromosomes in Leukemia. *Cancer Genet. Cytogenet.* **73:**1–7, 1994.
30. Yeyati, P. L., Shaknovich, R., Boterashvili, S., et al. Leukemia translocation protein PLZF inhibits cell growth and expression of cyclin A. *Oncogene* **18:**925–934, 1999.
31. Ruthardt, M., Orleth, A., Tomassoni, L., et al. The acute promyelocytic leukaemia specific PML and PLZF proteins localize to adjacent and functionally distinct nuclear bodies. *Oncogene* **16:**1945–1953, 1998.
32. Arnould, C., Philippe, C., Bourdon, V., et al. The signal transducer and activator of transcription STAT5b gene is a new partner of retinoic acid receptor alpha in acute promyelocytic-like leukaemia. *Hum. Mol. Genet.* **8:**1741–1749, 1999.
33. Downing, J. R. The AML1-ETO chimaeric transcription factor in acute myeloid leukaemia: biology and clinical significance. *Br. J. Haematol.* **106:**296–308, 1999.
34. Downing, J. R. The core-binding factor leukemias: lesson learned from murine models. *Curr. Opin. Genet. Dev.* **13:**48–54, 2003.
35. Yuan, Y., Zhou, L., Miyamoto, T., et al. AML1-ETO expression is directly involved in the development of acute myeloid leukemia in the presence of additional mutations. *Proc. Natl. Acad. Sci. USA* **98:**10398–10403, 2001.
36. Higuchi, M., O'Brien, D., Kumaravelu, P., et al. Expression of a conditional AML1-ETO oncogene bypasses embryonic lethality and establishes a murine model of human t(8;21) acute myeloid leukemia. *Cancer Cell* **1:**63–74, 2002.
37. Klampfer, L., Zhang, J., Zelenetz, A. O., Uchida, H., and Nimer, S. D. The AML1/ETO fusion protein activates transcription of BCL-2. *Proc. Natl. Acad. Sci. USA* **93:**14,059–14,064, 1996.
38. Linggi, B., Muller-Tidow, C., van de Locht, L., et al. The t(8;21) fusion protein, AML1 ETO, specifically represses the transcription of the p14(ARF) tumor suppressor in acute myeloid leukemia. *Nat. Med.* **8:**743–750, 2002.
39. Melnick, A., Carlile, G. W., McConnell, M. J., et al. AML-1/ETO fusion protein is a dominant negative inhibitor of transcriptional repression by the promyelocytic leukemia zinc finger protein. *Blood* **96:**3939–3947, 2000.
40. Tanaka, K., Tanaka, T., Kurokawa, M., et al. The AML1/ETO(MTG8) and AML1/Evi-1 leukemia-associated chimeric oncoproteins accumulate PEBP2beta(CBFbeta) in the nucleus more efficiently than wild-type AML1. *Blood* **91:**1688–1699, 1998.
41. Mitterbauer, M., Kusec, R., Schwarzinger, I., Haas, O. A., Lechner, K., and Jaeger, U. Comparison of karyotype analysis and RT-PCR for AML1/ETO in 204 unselected patients with AML. *Ann. Hematol.* **76:**139–143, 1998.
42. Le Beau, M. M., Larson, R. A., Bitter, M. A., Vardiman, J. W., Golomb, H. M., and Rowley, J. D. Association of an inversion of chromosome 16 with abnormal marrow eosinophils in acute myelomonocytic leukemia. A unique cytogenetic- clinicopathological association. *N. Engl. J. Med.* **309:**630–636, 1983.
43. Sun, X., Medeiros, L. J., Lu, D., et al. Dysplasia and high proliferation rate are common in acute myeloid leukemia with inv(16)(p13q22). *Am. J. Clin. Pathol.* **120:**236–245, 2003.
44. Haferlach, T., Winkemann, M., Loffler, H., et al. The abnormal eosinophils are part of the leukemic cell population in acute myelomonocytic leukemia with abnormal eosinophils (AML-M4Eo) and carry the pericentric inversion 16: a combination of May–Grunwald–Giemsa staining and fluorescence in situ hybridization. *Blood* **87:**2459–2463, 1996.
45. Liu, P. P., Wijmenga, C., Hajra, A., et al. Identification of the chimeric protein product of the CBFB-MYH11 fusion gene in inv(16) leukemia cells. *Genes Chromosomes Cancer* **16:**77–87, 1996.
46. Delaunay, J., Vey, N., Leblanc, T., et al. Prognosis of inv(16)/t(16;16) acute myeloid leukemia (AML): a survey of 110 cases from the French AML Intergroup. *Blood* **102:**462–469, 2003.
47. Kundu, M., and Liu, P. P. Function of the inv(16) fusion gene CBFB-MYH11. *Curr. Opin. Hematol.* **8:**201–205, 2001.
48. Mancini, M., Cedrone, M., Diverio, D., et al. Use of dual-color interphase FISH for the detection of inv(16) in acute myeloid leukemia at diagnosis, relapse and during follow-up: a study of 23 patients. *Leukemia* **14:**364–368, 2000.
49. Viswanatha, D. S., Chen, I., Liu, P. P., et al. Characterization and use of an antibody detecting the CBFbeta-SMMHC fusion protein in inv(16)/t(16;16)-associated acute myeloid leukemias. *Blood* **91:**1882–1890, 1998.
50. Schoch, C., Schnittger, S., Klaus, M., et al. AML with 11q23/MLL abnormalities as defined by the WHO classification: incidence, partner chromosomes, FAB subtype, age distribution, and prognostic impact in an unselected series of 1897 cytogentcially analyzed AML cases. *Blood* **102:**2395–2402, 2003.
51. Poppe, B., Vandesompele, J., Schoch, C., et al. Expression analyses identify MLL as a prominent target of 11q23 amplification and support an etiologic role for MLL gain of function in myeloid malignancies. *Blood* **103:**229–235, 2004.
52. Thirman, M. J., Gill, H. J., Burnett, R. C., et al. Rearrangement of the MLL gene in acute lymphoblastic and acute myeloid leukemias with 11q23 chromosomal translocations. *N. Engl. J. Med.* **329:** 909–914, 1993.
53. Ferrando, A. A., Armstrong, S. A., Neuberg, D. S., et al. Gene expression signatures in MLL-rearranged T-lineage and B-precursor acute leukemias: dominance of HOX dysregulation. *Blood* **102:** 262–268, 2003.
54. Repp, R., Borkhardt, A., Haupt, E., et al. Detection of four different 11q23 chromosomal abnormalities by multiplex-PCR and fluorescence-based automatic DNA-fragment analysis. *Leukemia* **9:**210–215, 1995.
55. Soekarman, D., von Lindern, M., Daenen, S., et al. The translocation (6;9)(p23;q34) shows consistent rearrangement of two genes and defines a myeloproliferative disorder with specific clinical features. *Blood* **79:**2990–2997, 1992.
56. Oyarzo, M., Lin, P., Glassman, A., et al. Acute myeloid leukemia with t(6;9)(p23;q34) is associated with dysplasia and high frequency of flt3 gene mutations. *Am. J. Clin. Pathol.* **122:**348–388, 2004.
57. von Lindern, M., Fornerod, M., van Baal, S., et al. The translocation (6;9) associated with specific subtype of acute myeloid leukemia, results in fusion of two genes, *dek* and *can*, and the expression of a chimeric, leukemia-specific *dek-can* mRNA. *Mol. Cell. Biol.* **12:**1687–1697, 1992.
58. Kraemer, D., Wozniak, R. W., Blobel, G., et al. The human CAN protein, a putative oncogene product associated with myeloid leukemogenesis, is a nuclear pore complex protein that faces the cytoplasm. *Proc. Natl. Acad. Sci. USA* **91:**1519–1523, 1994.
59. Fornerod, M., Boer, J., van Baal, S., et al. Relocation of the carboxyterminal part of *CAN* from the nuclear envelope to the nucleus as a result of leukemia-specific chromosome rearrangements. *Oncogene* **10:**1739–1748, 1995.
60. Greene, M. E., Mundschau, G., Wechsler, J., et al. Mutations in *GATA1* in both transient myeloproliferative disorder and acute megakaryoblastic leukemia of Down syndrome. *Blood Cells Mol. Dis.* **31:**351–356, 2003.
61. Ma, Z., Morris, S. W., Valentine, V., et al. Fusion of two novel genes, RBM15 and MKL1, in the t(1;22)(p13;q13) of acute megakaryoblastic leukemia. Nat. Genet. **28:**220–221, 2001.
62. Mercher, T., Busson-Le Coniat, M., Khac, F. N., et al. Recurrence of *OTT-MAL* fusion in t(1;22) of infant AML-M7. *Genes Chromosomes Cancer* **33:**22–28, 2002.

63. Kurzrock, R., Kantarjian, H. M., Druker, B. J., et al. Philadelphia chromosome-positive leukemias: from basic mechanisms to molecular therapeutics. *Ann. Intern. Med.* **138:**819–830, 2003.
64. Dewald, G. W. Interphase FISH studies of chronic myeloid leukemia. *Methods Mol. Biol.* **204:**311–342, 2002.
65. McWhirter, J. and Wang, J. An actin binding function contributes to transformation by the Bcr-Abl oncoprotein of Philadelphia chromosome-positive human leukemias. EMBO J. **12:**1533–1546, 1993.
66. Wetzler, M., Talpaz, M., Van Etten, R. A., et al. Subcellular localization of Bcr, Abl, and Bcr-Abl proteins in normal and leukemic cells and correlation of expression with myeloid differentiation. *J. Clin. Invest.* **92:**1925–1939, 1993.
67. Sattler, M. and Griffin, J. D. Molecular mechanisms of transformation by the BCR-ABL oncogene. Semin. Hematol. **40(Suppl. 2):** 4–10, 2003.
68. Pane, F., Frigeri, F., Sindona, M., et al. Neutrophilic-chronic myeloid leukemia: a distinct disease with a specific molecular marker (BCR/ABL with C3/A2 junction). *Blood* **88:**2410–2414, 1996.
69. Huntly, B. J., Bench, A., and Green, A. R. Double jeopardy from a single translocation: deletions of the derivative chromosome 9 in chronic myeloid leukemia. *Blood* **102:**1160–1168, 2003.
70. Yamaguchi, H., Inokuchi, K., Sakuma, Y., et al. Mutation of the p51/p63 gene is associated with blastic crisis in chronic myelogenous leukemia. *Leukemia* **15:**1729–1734, 2001.
71. Hernandez-Boluda, J. C., Cervantes, F., Colomer, D., et al. Genomic p16 abnormalities in the progression of chronic myeloid leukemia into blast crisis: a sequential study in 42 patients. *Exp. Hematol* **31:**204–210, 2003.
72. Guo, J. Q., Lin, H., Kantarjian, H., et al. Comparison of competitive-nested PCR and real-time PCR in detecting BCR-ABL fusion transcripts in chronic myeloid leukemia patients. *Leukemia* **16:**2447–2453, 2002.
73. Druker, B. J., Talpaz, M., Resta, D. J., et al. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N. Engl. J. Med.* **344:**1031–1037, 2001.
74. Kantarjian, H., Sawyers, C., Hochhaus, A., et al. Hematologic and cytogenetic responses to imatinib mesylate in chronic myelogenous leukemia. *N. Engl. J. Med.* **346:**645–652, 2002.
75. Druker, B. J., Sawyers, C. L., Kantarjian, H., et al. Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome. *N. Engl. J. Med.* **344:**1038–1042, 2001.
76. Gorre, M. E., Mohammed, M., Ellwood, K., et al. Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. *Science* **293:**876–880, 2001.
77. Branford, S., Rudzki, Z., Walsh, S., et al. Detection of BCR-ABL mutations in patients with CML treated with imatinib is virtually always accompanied by clinical resistance, and mutations in the ATP phosphate-binding loop (P-loop) are associated with a poor prognosis. *Blood* **102:**276–283, 2003.
78. Golub, T. R., Barker, G. F., Lovett, M., et al. Fusion of PDGF receptor beta to a novel *ets*-like gene, *tel*, in chronic myelomonocytic leukemia with t(5;12) chromosome translocation. *Cell* **77:**307–316, 1994.
79. Apperley, J. F., Gardembas, M., Melo, J. V., et al. Response to imatinib mesylate in patients with chronic myeloproliferative diseases with rearrangements of the platelet-derived growth factor receptor beta. *N. Engl. J. Med.* **347:**481–487, 2002.
80. Steer, E. J. and Cross, N. C. P. Myeloproliferative disorders with translocations of chromosome 5q31-35: role of platelet derived growth factor receptor beta. *Acta Haematol.* **107:**113–122, 2002.
81. Carroll, M., Ohno-Jones, S., Tamura, S., et al. CGP 57148, a tyrosine kinase inhibitor, inhibits the growth of cells expressing BCR-ABL, TEL-ABL, and TEL-PDGFR fusion proteins. *Blood* **90:** 4947–4952, 1997.
82. Magnusson, M. K., Meade, K. E., Nakamura, R., et al. Activity of STI571 in chronic myelomonocytic leukemia with a platelet-derived growth factor b receptor fusion oncogene. *Blood* **100:** 1088–1091, 2002.
83. Ross, T. S., Bernard, O. A., Berger, R., et al. Fusion of Huntington interacting protein 1 to platelet derived growth factor beta receptor (PDGFbetaR) in chronic myelomonocytic leukemia with t(5;7)(q33;q11.2). *Blood* **91:**4419–4426, 1998.
84. Schwaller, J., Anastasiadou, E., Cain, D., et al. H4 (D10S170), a gene frequently rearranged in papillary thyroid carcinoma, is fused to the platelet-derived growth factor receptor beta gene in atypical chronic myeloid leukemia with t(5;10)(q33;q22). *Blood* **97:**3910–3918, 2001.
85. Abe, A., Emi, N., Tanimoto, M., et al. Fusion of the platelet-derived growth factor receptor beta to a novel gene CEV14 in acute myelogenous leukemia after clonal evolution. *Blood* **90:**4271–4277, 1997.
86. Cogswell, P. C., Morgan, R., Dunn, M., et al. Mutations of the *ras* protooncogenes in chronic myelogenous leukemia: a high frequency of *ras* mutations in bcr/abl rearrangement-negative chronic myelogenous leukemia. *Blood* **74:**2629–2933, 1989.
87. Baxter, E. J., Kulkarni, S., Vizmanos, J. L., et al. Novel translocations that disrupt the platelet-derived growth factor receptor β (PDGFβ) gene in BCR-ABL-negative chronic myeloproliferative disorders. *Br. J. Haematol.* **120:**251–256, 2003.
88. Cools, J., DeAngelo, D. J., Gotlib, J., et al. A tyrosine kinase created by fusion of the *PDGFR*A and *FIP1L*1 genes as a therapeutic target of imatinib in idiopathic hypereosinophilic syndrome. *N. Engl. J. Med.* **348:**1201–1214, 2003.
89. Harrison, C. J. and Foroni, L. Cytogenetics and molecular genetics of acute lymphoblastic leukemia. *Rev. Clin. Exp. Hematol.* **6:**91–113, 2002.
90. Ferrando, A. A. and Look, A. T. Gene expression profiling in T-cell acute lymphoblastic leukemia. *Semin. Hematol.* **40:**274–280, 2003
91. Nashed, A. L., Rao, K. W., and Gulley, M. L. Clinical applications of BCR-ABL molecular testing in acute leukemia. *J. Mol. Diagn.* **5:**63–72, 2003.
92. Onciu, M., Bueso-Ramos, C., Medeiros, L. J., et al. Acute lymphoblastic leukemia in elderly patients: the Philadelphia chromosome may not be a significant adverse prognostic factor. *Am. J. Clin. Pathol.* **117:**716–720, 2002.
93. Loh, M. L. and Rubnitz, J. E. *TEL/AML1*-positive pediatric leukemia: prognostic significance and therapeutic approaches. *Curr. Opin. Hematol.* **9:**345–352, 2002.
94. Ford, A. M., Bennett, C. A., Price, C. M., et al. Fetal origins of the *TEL-AML1* fusion gene in identical twins with leukemia. *Proc. Natl. Acad. Sci. USA* **95:**4584–4588, 1998.
95. Hunger, S. Chromosomal translocations involving the E2A gene in acute lymphoblastic leukemia: clinical features and molecular pathogenesis. *Blood* **87:** 1211–1224, 1996.
96. Boomer, T., Varella-Garcia, M., McGavran, L., et al. Detection of E2A translocations in leukemias via fluorescence in situ hybridization. *Leukemia* **15:**95–102, 2001.
97. Haidar, M. A., Kantarjian, H., Manshouri, L., et al. ATM gene deletion in patients with adult acute lymphoblastic leukemia. *Cancer* **88:**1057–1062, 2000.
98. Kawamura, M., Kikuchi, A., Kobayashi, S., et al. Mutations of the p53 and ras genes in childhood t(1;19)-acute lymphoblastic leukemia. *Blood* **85:**2546–2552, 1995.
99. Lubbert, M., Mirro, J., Miller, C. W., et al. N-Ras gene point mutations in childhood acute lymphocytic leukemia correlate with a poor prognosis. *Blood* **75:**1163-1169, 1990.
100. Kikuchi, A., Hayashi, Y., Kobayashi, S., et al. Clinical significance of *TAL1* gene alteration in childhood T-cell acute lymphoblastic leukemia and lymphoma. *Leukemia* **7:**933–938, 1993.
101. Ferrando, A. A., Herblot, S., Palomero, T., et al. Biallelic transcriptional activation of oncogenetic transcription factors in T-cell acute lymphoblastic leukemia. *Blood* **103:**1909–1911, 2004.
102. Begley, C. G. and Green, A. R. The SCL gene: from case report to critical hematopoietic regulator. *Blood* **93:**2760–2770, 1999.
103. O'Neil, J., Ventura, J. J., Cusson, N., and Kelliher, M. NF-κB activation in premalignant mouse tal-1/scl thymocytes and tumors. *Blood* **102:**2593–2596, 2003.
104. Bernard, O. A., Busson-Le Coniat, M., Ballerini, P., et al. A new recurrent and specific cryptic translocation, t(5;14)(q35;q32), is associated with expression of the Hox11L2 gene in T acute lymphoblastic leukemia. *Leukemia* **15:**1495–1504, 2001.

105. Berger, R., Dastugue, N., Busson, M., et al. t(5;14)/HOXL2-positive T-cell acute lymphoblastic leukemia. A collaborative study of the Groupe Francais de Cytogenetique Hematologique (GFCH). *Leukemia* **17:**1851-1857, 2003.
106. Schneider, N. R., Carroll, A. J., and Shuster, J. J**.** New recurring cytogenetic abnormalities and association of blast cell karyotypes with prognosis in childhood T-cell acute lymphoblastic leukemia: a pediatric oncology group report of 343 cases. *Blood* **96:**2543–2549, 2000.
107. Murphy, S., Raimondi, S., Rivera, G., et al. Nonrandom abnormalities of chromosome 9p in childhood acute lymphoblastic leukemia: association with high-risk clinical features. *Blood* **74:**409–415, 1989.
108. Okuda, T., Shurtleff, S., Valentine, M., et al. Frequent deletion of $p16^{INK4A}$/MTS1 and $p15^{INK4B}$/MTS2 in pediatric acute lymphoblastic leukemia. *Blood* **85:**2321–2330, 1995.
109. Johansson, B., Moorman, A. V., Haas, O. A., et al. Hematologic malignancies with t(4;11)(q21;q23)—a cytogenetic, morphologic, immunophenotypic, and clinical study of 183 cases. European 11q23 Workshop participants. *Leukemia* **12:**779–787, 1998.
110. Tsujioka, T., Wada, H., Yamamori, S., et al. MLL/AF-1p fusion in therapy-related early pre-B acute lymphoblastic leukemia with t(1;11)(p32;q23) translocation developing in the relapse phase of acute promyelocytic leukemia. *Int. J. Hematol.* **78:**439–442, 2003.
111. Hiwatari, M., Taki, T., Taketani, T., et al. Fusion of an AF4-related gene, LAF4, to MLL in childhood acute lymphoblastic leukemia with t(2;11)(q11;q23). *Oncogene* **22:**2851–2855, 2003.
112. Kuefer, M. U., Chinwalla, V., Zeleznik-Le, N. J., et al. Characterization of the MLL partner gene AF15q14 involved in t(11;15)(q23;q14). *Oncogene* **22:**1418–1424, 2003.
113. Buhmann, R., Kurzeder, C., Rehklau, J., et al. CD40L stimulation enhances the ability of conventional metaphase cytogenetics to detect chromosome aberrations in B-cell chronic lymphocytic leukaemia cells. *Br. J. Haematol.* **118:**968–975, 2002.
114. Stilgenbauer, S., Lichter, P., and Dohner, H. Genetic features of B-cell chronic lymphocytic leukemia. *Rev. Clin. Exp. Hematol.* **4:**48–72, 2000.
115. Dewald, G. W., Brockman, S. R., Paternoster, S. F., et al. Chromosome anomalies detected by interphase fluorescence *in situ* hybridization: correlation with significant biological features of B-cell chronic lymphocytic leukemia. *Br. J. Haematol.* **121:**287–295, 2003.
116. Adachi, M., Tefferi, A., Greipp, P. R., et al. Preferential linkage of *bcl*-2 to immunoglobulin light chain gene in chronic lymphocytic leukemia. *J. Exp. Med.* **171:**559–564, 1990.
117. Michaux, L., Mecucci, C., Stul, M., et al. BCL3 rearrangements and t(14;19)(q32;q13) in lymphoproliferative disorders. *Genes Chromosomes Cancer* **15:**38–47, 1996.
118. McKeithan, T. W., Takimoto, G. S., Ohno, H., et al. BCL3 rearrangements and t(14;19) in chronic lymphocytic leukemia and other B-cell malignancies: a molecular and cytogenetic study. *Genes Chromosomes Cancer* **20:**64–72, 1997.
119. Satterwhite, E., Sonoki, T., Willis, T. G., et al. The BCL11 gene family: involvement of BCL11A in lymphoid malignancies. *Blood* **98:**3413–3420, 2001.
120. Kuppers, R., Sonoki, T., Satterwhite, E., et al. Lack of somatic hypermutation og IG V(H) genes in lymphoid malignancies with t(2;14)(p13;q32) translocaton involving the BCL11A gene. *Leukemia* **16:**937–939, 2002.
121. Medeiros, L. J., Van Krieken, J. H. J. M., Jaffe, E. S., et al. Association of bcl-1 rearrangements with lymphocytic lymphoma of intermediate differentiation. *Blood* **76:**2086–2090, 1990.
122. Damle, R. N., Wasil, T., Fais, F., et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood* **94:**1840–1847, 1999.
123. Hamblin, T. J., Davis, Z., Gardiner, A., et al. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood* **94:**1848–1854, 1999.
124. Oscier, D. G., Gardiner, A. C., Mould, S. J., et al. Multivariate analysis of prognostic factors in CLL: clinical stage, *IGVH* gene mutational status, and loss or mutation of the *p53* gene are independent prognostic factors. *Blood* **100:**1177–1184, 2002.
125. Klein, U., Tu, Y., Stolovitzky, G. A., et al. Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. *J. Exp. Med.* **194:**1625–1638, 2001.
126. Korz, C., Pscherer, A., Benner, A., et al. Evidence for distinct pathomechanisms in B-cell chronic lymphocytic leukemia and mantle cell lymphoma by quantitative expression analysis of cell cycle and apoptosis-associated genes. *Blood* **99:**4554–4561, 2002.
127. Foon, K. A., Thiruvengadam, R., Saven, A., et al. Genetic relatedness of lymphoid malignancies. Transformation of chronic lymphocytic leukemia as a model. *Ann. Intern. Med.* **119:**63–73, 1993.
128. Bea, S., Lopez-Guillermo, A., Ribas, M., et al. Genetic imbalances in progressed B-cell chronic lymphocytic leukemia and transformed large-cell lymphoma (Richter's syndrome). *Am. J. Pathol.* **161:**957–968, 2002.
129. Schlette, E., Bueso-Ramos, C., Giles, F., et al. Mature B-cell leukemias with more than 55% prolymphocytes. A heterogeneous group that includes an unusual variant of mantle cell lymphoma. *Am. J. Clin. Pathol.* **115:**571–581, 2001.
130. Lens, D., Dyer, M. J., Garcia-Marco, J. M., et al. p53 abnormalities in CLL are associated with excess prolymphocytes and poor prognosis. *Br. J. Haematol.* **99:**848–857, 1997.
131. Merchant, S., Schlette, E., Sanger, W., et al. Mature B-cell leukemias with more than 55% prolymphocytes: report of 2 cases with Burkitt lymphoma-type chromosomal translocations involving c-myc. *Arch. Pathol. Lab. Med.* **127:**305–309, 2003.
132. Lens, D., De Schouwer, P. J., Hamoudi, R. A., et al. p53 abnormalities in B-cell prolymphocytic leukemia. *Blood* **89:**2015–2023, 1997.
133. Garand, R., Goasguen, J., Brizrad, A., et al. Indolent course as a relatively frequent presentation in T-prolymphocytic leukemia. Groupe Francais d'Hematologie Cellulaire. *Br. J. Haematol.* **103:**488–494, 1998.
134. Pekarsky, Y., Hallas, C., and Croce, C. M. Molecular basis of mature T-cell leukemia. *JAMA* **286:**2308–2314, 2001.
135. Boultwood, J. Ataxia telangiectasia gene mutations in leukaemia and lymphoma. *J. Clin. Pathol.* **54:**512–516, 2001.

VIII Applications of Molecular Diagnosis for Infectious Diseases

34 Molecular Testing for *Chlamydia trachomatis* and *Neisseria gonorrhoeae*

ALEXANDRA VALSAMAKIS

1. INTRODUCTION

Chlamydia trachomatis and *Neisseria gonorrhoeae* are currently the two most prevalent pathogens in the United States. The costs of infections, to individuals and society, are significant. Unfortunately, these pathogens can be difficult to detect with conventional diagnostic methods. New molecular techniques have improved diagnosis and have made screening possible. This aim of this chapter is to provide practical information on molecular testing for *C. trachomatis* and *N. gonorrhoeae* that is useful to practitioners of molecular pathology. Introductory information on the clinical significance and conventional diagnostic tests is provided. Issues confronting molecular testing for these pathogens are discussed. Descriptions of each molecular platform are provided, including performance characteristics and other parameters of functionality.

2. CHLAMYDIA TRACHOMATIS

2.1. CLINICAL BACKGROUND *Chlamydia* trachomatis is the most common reportable communicable infection in the United States. Reported rates have risen fivefold (51–278 cases/100,000 persons) in the past 20 yr, likely the result of the combination of better diagnostic tests, implementation of screening, and increased infection rates *(1)*. The primary route of transmission is through sexual contact. Intrapartum transmission can produce neonatal disease. Rates of infection are highest in adolescents (ages 15–19) and young adults (ages 20–24) *(2)*. Most infections are asymptomatic, resulting in a large reservoir for transmission.

Infection can cause urethritis, endometritis, salpingitis, pelvic inflammatory disease (PID), and perihepatitis in women and urethritis, epididymitis, and proctitis in men. The most common signs of infection are mucopurulent cervicitis in women and urethritis in men. Significant damage to reproductive organs can occur despite the absence of symptoms, resulting in infertility. Fallopian tube damage, from PID or salpingitis, causes infertility and predisposition to ectopic pregnancy *(3)*. In pregnant women, *C. trachomatis* infection is associated with preterm birth *(4)* and other complications such as stillbirth, premature rupture of membranes, and postpartum endometritis *(5)*. *C. trachomatis* infections can have long term consequences in men, including infertility, urethral strictures, and chronic prostatitis. Conjunctivitis and pneumonia can occur in neonates as a result of perinatal acquisition from mothers with cervicitis.

The incidence of asymptomatic infection, its impact on individuals, and its influence on the prevalence of disease in the community has led multiple professional organizations (national and international) and federal agencies to recommend that all sexually active women aged 25 yr and younger and all other asymptomatic women at risk of infection be screened for *C. trachomatis* genital infection *(6–9)*. Screening and treatment of asymptomatic pregnant women at risk of infection has also been recommended because diagnosis and treatment can improve pregnancy outcomes. Screening can prevent pelvic inflammatory disease *(10)* and is cost-effective. A meta-analysis of multiple cost-effectiveness studies concluded that screening urine with molecular assays was cost-effective when prevalence was 3.1–10% and cost saving when prevalence was as low as 1.1% *(11)*. Annual screening has been recommended, however, more frequent testing might be warranted because reinfection has been found to occur every 7 mo in sexually active women <25 yr of age, despite adequate treatment *(12)*. No direct evidence for decreased rates of infection in women has been demonstrated as a result of screening for *C. trachomatis* infections in men; therefore, screening for infection in men is not currently recommended.

Chlamydia trachomatis is an obligate intracellular organism with a unique life cycle among bacteriaciae. An infectious, metabolically inactive form (elementary body [EB]) enters host cells. These reorganize into more loosely configured reticulate bodies (RBs) that are noninfectious and metabolically active. RBs divide by binary fission and asynchronously condense into EBs. Infected cells lyse, EBs are released, and the infectious cycle repeats.

2.2. CONVENTIONAL DETECTION ASSAYS Conventional assays utilize either species-specific monoclonal antibodies against major outer membrane protein (MOMP) or

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

Table 1
Tests Recommended for *C. trachomatis/N. gonorrhoeae* Screening[a]

Pathogen	*Gender*	*Test*	*Specimen*	*Recommendation*
C. trachomatis	Female	NAAT	Endocervical swab or urine	Preferred test
		UNAHT	Endocervical swab	Acceptable test
		EIA	Endocervical swab	Suitable test
		DFA	Endocervical swab	Suitable test
		Culture	Endocervical swab	Suitable test
	Male	NAAT	Intraurethral swab or urine	Preferred test
		UNAHT	Intraurethral swab	Acceptable test
N. gonorrhoeae	Female	Culture	Endocervical swab	Preferred test: allows for antimicrobial susceptibility testing
		NAAT or UNAHT	Endocervical swab	Acceptable test if storage/transport conditions compromise culture sensitivity
		NAAT	Urine	Preferred test for non traditional collection settings
	Male	Culture	Intraurethral swab	Preferred test: allows for antimicrobial susceptibility testing
		NAAT or UNAHT	Intraurethral swab	Acceptable test if storage/transport conditions compromise culture sensitivity
		NAAT	Urine	Acceptable test

[a]United States Centers for Disease Control and Prevention recommendations *(1)*.
Abbreviations: NAAT, nucleic acid amplification test; UNAHT, unamplified nucleic acid hybridization test; EIA, enzyme immunoassay; DFA, direct fluorescent antigen.

genus-specific antibodies against lipopolysaccharide (LPS) to detect *C. trachomatis* growing in tissue culture cells or directly in specimens (cytologic smears or enzyme immunoassays). The major advantages of cell culture are its specificity and the ability to further characterize isolates. As a consequence, cell culture remains the method of choice when the results of diagnostic tests are to be used in cases of rape or sexual abuse *(1)*. The drawback of these assays compared to nucleic acid amplification tests (NAATs) is a relative lack of sensitivity. Additional disadvantages of cell culture include turnaround time and labor intensity. Tests that have been recommended for *C. trachomatis* genitourinary infection screening are listed in Table 1.

3. *NEISSERIA GONORRHOEAE*

3.1. CLINICAL BACKGROUND *Neisseria gonorrhoeae* is currently the second most common sexually transmitted microbe in the United States *(1)*. Although its overall incidence has decreased, gonorrhea continues to be problematic, most likely as a result of transmission by women who are largely asymptomatic. Individuals at greatest risk are inner-city residents less than 24 yr of age who have unprotected sex with multiple sexual partners *(13)*. Transmission requires direct, usually sexual contact and the bacterium is transmitted more effectively from men to women than from women to men. Intrapartum transmission can also occur, producing disease in neonates.

Neisseria gonorrhoeae causes symptomatic disease and asymptomatic infection. In men, gonorrheae manifests itself within 7 d of infection as acute urethritis, with a penile discharge that ranges from scant, clear to copious and mucopurulent. This is often accompanied by increased urinary frequency. Symptoms of women with acute endocervical infection include vaginal discharge and dysuria. The cervical os is friable and erythematous.

Anogenital and oropharyngeal infections also occur as a result of transmission during receptive oral/anal intercourse or by contamination with cervical secretions. Manifestations of anorectal disease include copious purulent rectal discharge, rectal pain, tenesmus, and bloody stools. Oropharyngeal infection can produce mild pharyngitis. Both anogenital and oropharyngeal infections can be asymptomatic.

Disseminated infection occurs in a minority of cases (1–3%). It is most commonly observed in untreated asymptomatic patients. Patients with complement deficiencies (C7, C8, C9) are particularly prone to dissemination. Syndromes caused by disseminated infection include dermatitis–arthritis syndrome, septic monoarticular arthritis, endocarditis, and meningitis. Organism is isolated from blood in less than 50% of cases.

Intrapartum transmission produces conjunctivitis in neonates (ophthalmia neonatorum). Symptoms include tearing, edema, and purulent exudates. Corneal scarring and perforation can occur if infection is untreated. This syndrome is uncommon in the developed world because of the administration of antibiotic eye drops at birth.

Asymptomatic infection can cause significant disease if untreated. It is also the greatest barrier to controlling transmission and disease. The estimated prevalence of asymptomatic infection is 5% in men in the general population *(14)*. Asymptomatic infections in men can cause epididymitis, prostatitis, and urethral strictures. In women, urethral abscesses, salpingitis, tubo-ovarian abscesses, endometritis, and Fitz–Hugh–Curtis syndrome (perihepatitis) are observed. Sequelae include infertility and ectopic pregnancy as a result of fallopian tube scarring and blockage. Complications during pregnancy include spontaneous abortion, premature rupture of membranes, premature labor, and chorioamnionitis *(13)*.

Gonorrhea screening has been recommended for reasons similar to those for *C. trachomatis* screening *(1,15,16)*. Testing should be performed on asymptomatic women at high risk of infection and on women at high risk during pregnancy. Screening of asymptomatic men is currently not recommended.

3.2. CONVENTIONAL DETECTION ASSAYS *Neisseria gonorrhoeae* is an aerobic Gram-negative diplococcus that can be isolated on enriched media (most are chocolate agar-based). Growth is stimulated by CO_2 and humidity. Some isolates have an obligate CO_2 requirement. Inoculation of plates at the bedside immediately after collection and transportation in a CO_2-containing environment results in maximal rates of successful isolation from genital sites. Plating and transport at low temperatures can inhibit growth because the organism is sensitive to temperature. The organism is most effectively grown from endocervical and urethral swab specimens. Culture is still the gold standard methodology for identification of the organism in nongenital specimens. Similar to *C. trachomatis*, culture is the diagnostic method of choice in cases of rape and sexual abuse. The advantages of culture for gonorrhea diagnosis are sensitivity, specificity, low cost, the suitability of multiple specimen types (pharyngeal and rectal in addition to urethral and cervical specimens), and the ability to retain the isolate for further study (antibiotic susceptibility testing, medicolegal purposes, typing for epidemiologic studies). The major disadvantages of culture are turnaround time (24–72 h for presumptive culture results) and the inability to screen for gonorrhea by noninvasive urine sampling. Gram stain of urethral exudates in symptomatic men has sensitivity and specificity comparable to culture *(17)*. In asymptomatic men, Gram stain of urethral specimens is not as sensitive as culture *(17)*. The use of Gram stain in the diagnosis of gonorrhea in women is not recommended because of test performance in endocervical specimens [low sensitivity compared to Gram stain of urethral specimens from symptomatic men and the requirement for skilled microscopist to achieve optimal specificity *(1)*]. Tests that have been recommended for screening for *N. gonorrhoeae* infection are listed in Table 1.

4. MOLECULAR ASSAYS FOR THE DETECTION OF *C. TRACHOMATIS* AND *N. GONORRHOEAE:* OVERVIEW AND CURRENT ISSUES

Two assay formats, nucleic acid hybridization and nucleic acid amplification, have been developed for the detection of these organisms. Compared to culture, the first-generation hybridization assay (Pace-2, Gen-Probe) demonstrated improved sensitivity for detection of *C. trachomatis* and *N. gonorrhoeae (18)*. Second-generation assays with even greater sensitivity have been developed *(1,19,20)*. These tests include one hybridization signal amplification assay (Digene Hybrid Capture II CT/GC) and a number of NAATs (Abbott LCx, Becton Dickinson BDProbeTec ET, Gen-Probe APTIMA, and Roche Amplicor).

Determination of molecular assay performance has been complicated by comparisons to less sensitive reference methodologies. This is particularly true for the detection of *C. trachomatis*, because of the relative insensitivity of culture. Culture has been found to be an acceptable gold standard for the determination of *N. gonorrhoeae* amplification assay sensitivity and specificity.

Discrepant analysis, using either alternative amplification targets or methodologies, has been applied to help distinguish between false positives and true positives identified by more sensitive methods. Concern has been raised that the performance of discrepant analysis only on potential false-positive results might lead to an overestimation of sensitivity without sufficient resolution of specificity *(21–24)*. An alternative approach has been to use reference standards other than culture, including specimen standards (multiple NAAT results from a single specimen) and an infected patient standard (NAAT results from multiple sites) to define true chlamydia infection. In women, the clinical sensitivity of an individual NAAT is likely to be overestimated when an endocervical specimen standard is employed *(25,26)* because isolated chlamydia infection of the urethra can occur *(27)*. It has therefore been recommended that, in women, specimen standards be used to assess performance characteristics of a test from a single site *(1)* and that the infected patient standard should be used in assessing the clinical performance of individual assays *(26)*. In men, the use of a specimen standard is adequate for assessing clinical performance because a single genitourinary site is usually sampled (urethra, with urine as a noninvasive substitute).

Despite these standardization issues, second-generation molecular assays for the detection of *N. gonorrhoeae* and *C. trachomatis* have been found to perform well in symptomatic and asymptomatic patients. Therefore, these tests can be used in screening for infection and diagnosis of disease. They offer many advantages over conventional detection techniques, including sensitivitiy, decreased turnaround time, the ability to detect the organisms if culture is either not readily available, or if specimen transport conditions (including temperature, duration, CO_2 availability) required for successful culture cannot be provided. Molecular assays are theoretically useful for the detection of vancomycin-susceptible *N. gonorrhoeae* isolates whose growth is suppressed by antibiotics in the selective media normally used to isolate the organism. These organisms were once highly prevalent in certain geographic regions, but they are now uncommon *(28)*. The disadvantages of molecular assays include cost, the requirement for skilled laboratory technologists, the need for specialized laboratory design to prevent contamination, and the lack of an available isolate on which to perform antibiotic susceptibility testing, particularly for *N. gonorrhoeae*.

Specimens that have been cleared for use in most NAATs include (1) urine from men and women, (2) urethral swabs in men and women, and (3) endocervical swabs. In general, assay sensitivities in urine are lower than in swab specimens *(25,29–31)*. Test performance using self-collected low vaginal swabs (SCLVSs) has been studied for a variety of NAATs, and sensitivities comparable to other specimens have been demonstrated *(32–35)*. These specimens are relatively well accepted by women [better than endocervical swab, but less preferred compared to urine *(36–40)*]. One of the primary advantages of SCLVSs compared to urine is the lack of a requirement for storage at 4°C after collection, facilitating screening at remote sites. Unfortunately, SCLVSs have not yet been cleared for use in most NAATs. The ability to test relatively noninvasive specimens such as urine and SCLVSs should greatly enhance screening for *C. trachomatis* and *N. gonorrhoeae* infections.

Extragenital specimens have also not been cleared for use in NAATs. However, recent studies demonstrate that NAATs can effectively detect *C. trachomatis* and *N. gonorrhoeae* in

Table 2
***C. trachomatis* Molecular Assay Characteristics**

Test	*Manufacturer*	*Molecular method*	*Target*	*Cleared specimens*
LCx	Abbott	Ligase chain reaction	Multicopy cryptic plasmid	Assay no longer marketed in the United States
ProbeTec	Becton Dickinson	Strand displacement amplification	Multicopy cryptic plasmid	Endocervical swab, intraurethral swab (men), urine (men and women); symptomatic and asymptomatic patients
HCII CT/GC	Digene	Hybridization with signal amplification	39kb (4%) genomic sequence	Endocervical swab
HCII CT-ID	Digene		Cryptic plasmid	
Pace 2/Pace 2C	Gen-Probe	Unamplified nucleic acid hybridization	23s rRNA	Endocervical swab, intraurethral swab, conjunctiva
APTIMA Combo 2	Gen-Probe	Transcription mediated amplification	23s rRNA	Endocervical swab, vaginal swab, intraurethral swab (men), urine (men and women); symptomatic and asymptomatic patients
APTIMA CT	Gen-Probe	Transcription mediated amplification	16s rRNA	Assay not yet cleared
AMPLICOR CT/NG	Roche Diagnostics	Polymerase chain reaction	Multicopy cryptic plasmid	Endocervical swab, intraurethral swab (men), urine (men and women); symptomatic and asymptomatic patients

oropharyngeal and rectal specimens *(41–47)*. Initial studies have demonstrated that the prevalence of pharyngeal and rectal gonorrhea is higher after NAAT testing than after culture *(43)*. These data suggest that NAATs might have a useful role in screening for extragenital infection and diagnosis of disease in high-risk populations such as men who have sex with men.

Confirmatory testing has been advocated for positive NAAT results since the use of these highly sensitive assays as screening tests might result in a significant number of false-positive results *(1)*. The proportion of positive results that are true positive (or positive predictive value [PPV]) is likely to be particularly low in populations with a low prevalence of infection. It has therefore been recommended that routine confirmatory testing should be considered when PPV is less than 90% *(1)*. Despite this recommendation, no multicenter trials have been performed to establish the utility of a specific test algorithm in the screening and confirmation of infection. Theoretically, the best confirmatory approach will use a different assay format to test an alternative specimen, procured either at the time of the initial specimen or during a follow-up visit, when the presumptive positive result is discussed. The impracticality of collecting, storing, and retrieving alternate specimens limits the feasibility of this algorithm. The use of less sensitive non-NAATs as confirmatory tests is not effective because of the potential for false-negative confirmatory test results. A two-step algorithm comprised of a nonamplified nucleic acid hybridization assay with a modified cutoff followed by confirmatory NAAT testing has been proposed *(1)*. A revised cutoff in the indeterminate or gray zone below the standard value would increase the sensitivity of the nonamplified nucleic acid hybridization assay. Currently, laboratories seeking to adopt this approach would need to verify appropriate cutoffs because a revised value is not provided by the manufacturer. The most straightforward solution to confirmation will be a combination of correct application of selective screening criteria by health care providers and the development of confirmatory assays that can amplify alternate targets using both the original specimen and NAAT screening platform.

5. COMMERCIAL MOLECULAR ASSAYS FOR THE DETECTION OF *C. TRACHOMATIS* AND *N. GONORRRHOEAE*

Information on individual assays, including manufacturer, method, format, target, and cleared specimens, is described in Tables 2 and 3. Additional features of each assay (platform design, sensitivity, specificity, collection information, transport/storage requirements, performance on non-cleared specimens, controls, reporting issues, and throughput) are discussed below. Assays are listed alphabetically according to manufacturer.

5.1. LCX The sensitivity/specificity of this assay for the detection of *C. trachomatis* is 81–100%/97.5–100% in endocervical swabs *(25,48)* and 98–100%/99.3–100% in male urethral swabs *(48)*. Sensitivity/specificity in urine is 69.6–96.4%/99.6–100% in women and 93.5–96.4%/ 99.8–100% in men *(48)*. For *N. gonorrhoeae* detection, the sensitivity/specificity of LCx is 89.5–97.4%/99.1–100% in endocervical swabs and 95.9–100/99.8–100% in male urethral swabs *(49–52)*. Sensitivity/specificity in urine is 50–94.7%/ 100% in women and 88.9–98%/99.6–100% in men *(50–52)*. This assay has been used extensively in screening programs and in clinical trials; however, it is no longer available in the United States because of inconsistencies in manufacturing.

5.2. PROBETEC ET CHLAMYDIA TRACHOMATIS AND NEISSERIA GONORRHOEAE AMPLIFIED DNA ASSAYS In this platform, detection of *C. trachomatis* and *N. gonorrhoeae* requires testing for each pathogen in separate reaction

Table 3
***N. gonorrhoeae* Molecular Assay Characteristics**

Test	*Manufacturer*	*Molecular Method*	*Target*	*Cleared specimens*
LCx	Abbott	Ligase chain reaction	Opa gene	Assay no longer marketed in the United States
ProbeTec	Becton Dickinson	Strand displacement amplification	Multicopy chromosomal pilin gene	Endocervical swab, intraurethral swab (men), urine (men and women); symptomatic and asymptomatic patients
HCII CT/GC	Digene	Hybridization with signal amplification	9.6 kb (0.5%) genomic sequence	Endocervical swab
HCII GC-ID	Digene		Cryptic plasmid	
Pace 2/Pace 2C	Gen-Probe	Unamplified nucleic acid hybridization	16s rRNA	Endocervical swab, intraurethral swab, conjunctiva
APTIMA Combo 2	Gen-Probe	Transcription-mediated amplification	16s rRNA	Endocervical swab, vaginal swab, intraurethral swab (men), urine (men and women); symptomatic and asymptomatic patients
APTIMA NG	Gen-Probe	Transcription-mediated amplification	16s rRNA	Assay not yet cleared
AMPLICOR CT/NG	Roche Diagnostics	Polymerase chain reaction	Cytosine methyltransferase gene homolog	Endocervical swab, intraurethral swab (men), urine (men); symptomatic and asymptomatic patients

tubes. Using an infected patient standard, the sensitivity/specificity of this assay for the detection of *C. trachomatis* is 92.8%/98.1% in endocervical swabs and 92.5%/96.4% in male urethral swabs *(30)*. In urine, screening studies combining data from men and women demonstrated sensitivities/specificities of 95.3–96%/99.3–100% *(53,54)*. In one study, using an infected patient standard, sensitivity of detection in women was lower than men (80.5% vs 93.1%) *(30)*. For *N. gonorrhoeae* detection using an infected patient standard, the sensitivity/specificity of ProbeTec is 96.6%/99.5% in endocervical swabs and 98.5%/96.5% in male urethral swabs *(30)*. In urine, high sensitivity and specificity were observed in a mixed - gender screening study *(54)*. Similar to *C. trachomatis*, lower sensitivities were observed when sensitivity and specificity were calculated separately for women (84.9%) and men (97.9%) *(30)*. High specificities were also observed in this study (99.4% in women, 98.5% in men). Analytical specificity studies have shown crossreactivity with other nongonococcal *Neisseria* species that rarely cause genital disease, including *N. flavescens*, *N. lactamica*, *N. subflava*, and *N. cinerea (55)*.

Collection devices cleared for use in this assay include only male and female swabs formulated by the manufacturer. Endocervical and urethral specimens are collected and transported on dry swabs. Swab specimens are stable for 4–6 days at 2–27°C. Urine specimens are stable for 4–6 d at 2–8°C and 2 d at 15–27°C. Vaginal swabs have been found to perform similarly to endocervical swabs *(33)* but are not cleared for use in the assay.

An optional amplification control to detect specimen-specific inhibition of strand displacement amplification is included. It is amplified in a separate reaction tube from *C. trachomatis* and *N. gonorrhoeae* tests. A gray zone of low positive MOTA ("method other than acceleration", the read-out for SDA assays) scores from 2000 to 9999 is described in the package insert. This gray zone is substantiated by a study that demonstrated 80.8% and 33.3% of *C. trachomatis* and *N. gonorrhoeae* results were reproduced in this range, compared to 96.7% for both pathogens at higher MOTA scores (≥10,000) *(56)*.

The throughput using manual specimen preparation is 150 *C. trachomatis* and *N. gonorrhoeae* detection results in an 8-h shift *(54)*. A fully automated high-throughput system (Viper) that provides 564 *C. trachomatis* results/8-h shift or 552 *C. trachomatis* and *N. gonorrhoeae* results/8-h shift is available.

5.3. HYBRID CAPTURE II CT/GC, CT-ID, AND GC-ID

This platform consists of an initial screening assay that detects *C. trachomatis /N. gonorrhoeae* and individual identification assays that are performed on specimens that are initially positive. The second panel of tests can function as a confirmatory platform. The reported sensitivity/specificity of Hybrid Capture II (HCII) for the detection of *C. trachomatis* and *N. gonorrhoeae* in endocervical specimens is 97.2–100%/ 98.2–99.8% and 92.2–100%/98.5–99.7%, respectively *(57–59)*. Assay-specific conical brushes and sample medium obtainable from the manufacturer are used to collect and transport endocervical specimens. Specimens can be shipped at room temperature and are stable when stored at room temperature for 14 d or at –20°C for up to 3 mo prior to testing.

This assay is not cleared for use with male urethral swabs; however, sensitivity/specificity of 95.6%/100% and 100%/99% for the detection of *C. trachomatis* and *N. gonorrhoeae* has been reported for this specimen *(60)*. No data on extragenital specimens have been reported. Controls for the presence of specimen-specific amplification inhibitors are not required because pathogens are detected via amplification of hybridization signal

rather than target amplification. HCII screening assay results are reported as positive or not detected, with no gray zone. Positive specimens are tested further with CT-ID and GC-ID tests that have equivocal zones for results. Repeat testing with a new specimen or testing the original specimen with an alternate assay is recommended for equivocal identification results. The throughput of the manual microwell plate assay is 96 results in 6–8 h. A high-throughput Rapid Capture System that tests 352 specimens/8-h shift (up to four 96-well plates) is available. This automated platform performs liquid and plate handling after an initial manual specimen denaturation step. Reported sensitivity/specificity for detection of *C. trachomatis* and *N. gonorrhoeae* in endocervical specimens by the Rapid Capture system is 95.6%/100% and 100%/99%, respectively *(61)*.

5.4. PACE 2/PACE 2C This platform consists of an initial screening assay that detects *C. trachomatis* /*N. gonorrhoeae* (Pace 2C) and individual identification assays that are performed on specimens that are initially positive (Pace 2 for the detection of *C. trachomatis* and *N. gonorrhoeae*). These first-generation molecular assays are less sensitive than newer amplification assays *(25,50)*. The sensitivity/specificity of Pace 2 for the detection of *C. trachomatis* in endocervical specimens is 75.5–95%/97.1–100% *(25,62–65)*. In women and men, the sensitivity/specificity for detection of *N. gonorrhoea* is 97.1–100%/99.1–99.6% *(66–68)*. Pace 2C performance characteristics are similar *(18)*. Assay-specific specimen collection devices and transport medium are required. Several studies compared the detection of *N. gonorrhoeae* by Pace 2 and culture in pharyngeal and rectal specimens *(46,69)*. Specimens are stable at room temperature for 7 d according to the package insert. This could represent a conservative estimate because *N. gonorrhoeae* has been detected after storage for 1 mo at room temperature *(65)*. Results are reported out as positive or not detected, with no equivocal zone. Competitive probe assays for confirmation of positive results are available. The Pace platform is the only assay approved for use with conjunctival specimens. Assay throughput is 100 tests/3 h.

Assay performance for the detection of *N. gonorrhoeae* in nongenital specimens has been examined. Sensitivity is lower in pharyngeal than in rectal specimens (71–87%) *(46,69)*. Test specificity was high in both sample types.

5.5. APTIMA COMBO 2 In this platform, chemiluminescent read-outs for *C. trachomatis* and *N. gonorrhoeae* follow rapid ("flasher") and slow ("glower") kinetics, respectively, allowing simultaneous detection in a single tube. Using an infected patient standard, the sensitivity/specificity of Aptima Combo 2 (AC2) for the detection of *C. trachomatis* and *N. gonorrhoeae* in endocervical specimens is 92.8–94.1%/97.6–99.4% and 98.6–99.2%/98.7–99.8%, respectively *(26,31)*. In urine from women, sensitivity/specificity of *C. trachomatis* and *N. gonorrhoeae* was also high (94.7%/98.9% and 91.3%/99.3%, respectively) *(31)*. The performance of this platform in men has not been reported in the literature other than as a study of assay performance in urine from a population of boys and girls undergoing screening in a high school setting (sensitivity/specificity 100%/98.8%, in the combined population using a specimen standard) *(53)*. According to the package insert, the sensitivity/specificity for *C. trachomatis* detection in male urethral swabs and urine is 95.5%/97.5% and 97.9%/98.5%, respectively. The sensitivity/specificity for *N. gonorrhoeae* detection in male urethral swabs and urine is 99.1%/97.8% and 98.5%/99.6%, respectively. Performance characteristics were defined using an infected patient standard.

Aptima Combo 2 is the only assay approved for *C. trachomatis* and *N. gonorrhoeae* detection in physician- or patient-collected vaginal swabs. There are no published reports of assay performance on nongenital specimens. Assay-specific specimen kits containing collection devices and transport medium are required for male urethral, endocervical, and vaginal swabs. Swabs collected for other diagnostic tests including LCx, culture (M4 medium), and DFA and stored at 4°C for up to 2 wk have been found to be suitable for AC2 testing *(70)*. The specimen collection tubes contain cell lysis and RNA protectant reagents. Specimens stored in collection tubes are stable for prolonged periods at room temperature (60 d at 2–30°C or 90 d at –20°C to –70°C for swabs; 30 d at 2–30°C for urine). Urines collected and transported in sterile cups are stable for 24 h at room temperature.

An internal control for amplification inhibition by individual specimens is not included in the AC2 assay. However, extraction of pathogen DNA by target capture could obviate the need for this. Results are automatically interpreted by luminometer software and reported out as positive, not detected, equivocal, or invalid. The package insert recommends that equivocal or invalid results be repeated, but protocols for further testing and result reporting are not specified. Individual assays for the detection of *C. trachomatis* and *N. gonorrhoeae* can be obtained as analyte-specific reagents. After performance characteristics are verified by the user, these tests can be implemented as confirmatory assays for AC2 because alternative targets are amplified (Table 1).

The throughput of AC2 is 100 tests/5 h or 200 tests/6 h, depending on assay configuration. A fully automated platform, with three levels of throughput (400, 800, or 1600 results/8-h shift) is available (Tigris).

5.6. AMPLICOR CT/NG In this platform, detection of *C. trachomatis* and *N. gonorrhoeae* requires testing for each pathogen in separate reaction tubes. In large multicenter trials using infected patient standards, the performance of the semi-automated platform COBAS Amplicor CT/NG for the detection of *C. trachomatis* was uniformly high in swab and urine specimens from symptomatic and asymptomatic women and men (>85% sensitivity/>98.4% specificity) *(29)*. Assay sensitivity for the detection of *N. gonorrhoeae* was more variable *(71)*. Low sensitivity was observed in urine (42.3% in asymptomatic men, 64.8% in women) and in urethral swabs from asymptomatic men (73.1%). Higher sensitivities were observed for other genital specimen types (>92.4%). Specificities were uniformly high (>98.8%). The manual microwell plate assay performed similarly. In analytical specificity studies, crossreactivity was observed with nongonoccocal *Neisseria* species that rarely cause genital disease (*N. flavescens*, *N. lactamica*, *N. sicca*) *(55)*.

This assay is cleared for use with many different swab types (dacron, rayon, calcium alginate, on plastic or nonaluminum

shafts) and transport media (2SP, Bartels ChlamTrans, Chlamydia Transport Medium [Bartels Inc.]; SPG CTM, M4 [MicroTest Inc.]). At room temperature, swabs and urines are stable for 1 h and 24 h, respectively. Swabs and urines can be stored at 2–8°C for 7 d and at –20°C for 30 d, respectively. Testing for *C. trachomatis* can be performed on urines stored at –20°C for up to 60 d.

This assay has not been cleared for use for *N. gonorrhoeae* detection in urine from women and in urethral swabs from asymptomatic men. It has also not been cleared for use with nongenital specimens. However, studies have shown good agreement between Amplicor CT/NG and other methods (culture and other molecular assays) for the detection of *C. trachomatis* in pharyngeal, rectal, and conjunctival specimens *(41,44,45,72)*.

Internal controls are provided to detect specimen-specific amplification inhibition. The internal controls are added to each specimen, coamplified with potential targets, and detected in separate reaction cups in the final steps of the *C. trachomatis* and *N. gonorrhoeae* detection assays. Both assays have equivocal zones for reporting, but ranges and recommendations for further testing differ for the detection of the two pathogens. Manual (96-well plate), semiautomated (COBAS), and automated platforms are available. The COBAS platforms require manual specimen processing and reaction setup. Amplification and detection are performed on the COBAS instrument. The throughput of the COBAS platform varies from 22 results in approx 3 h 45 min to 120 results in 10 h, depending on configuration. The automated platform (Tecan frontend/amplification on ABI 9600/Tecan backend) consists of front-end liquid handling and back-end plate handling/detection. Results for 288 specimens can be obtained in an 8-h shift.

6. FUTURE TRENDS IN *C. TRACHOMATIS* AND *N. GONORRHOEAE* MOLECULAR TESTING

The advent of molecular amplification assays has greatly improved the ability to detect these two pathogens. Along the way, numerous issues have been raised, but their resolution has likely resulted in an overall betterment of patient care. Hopefully, future surveillance studies will demonstrate a decline in prevalence, thereby substantiating efforts that have been expended to improve screening and treatment. It is likely that efforts to extend screening programs into nontraditional environments will continue to expand. The Internet, with its flexibility and enormous outreach, would be an ideal means to expand screening. Programs to distribute test results and recommendations for further care via the Web are currently in the planning stages. As the prevalence of these infections declines, the importance of confirmatory testing will rise proportionally. In the future, the availability of confirmatory assays for each of the above-listed platforms will be critical. Finally, new tests are evolving. A homogeneous (real time) polymerase chain reaction assay is currently in development. Platforms for testing for *C. trachomatis*, *N. gonorrhoeae*, and other sexually transmitted pathogens such as human papillomaviruses in single specimens will also likely evolve. Preliminary data on such platforms are available *(73)*. Current concerns with regard to combination testing include the potential for interspecimen contamination during processing and the lack of controls for specimen adequacy. However, it is likely that these hurdles will be cleared so that effective testing is available, as has been demonstrated throughout the evolution of this field.

REFERENCES

1. Johnson, R. E., Newhall, W. J., Papp, J. R., et al. Screening tests to detect *Chlamydia trachomatis* and *Neisseria gonorrhoeae infections*—2002. *MMWR Recomm. Rep.* **51:**1–38, 2002; quiz CE31–CE34.
2. Stamm, W. E., *Chlamydia trachomatis* Infections of the Adult, in Wasserheit JN, ed. *Sexually Transmitted Diseases,* 3rd ed., J. N. Wasserheit, ed., McGraw-Hill, New York, pp. 407–422, 1999.
3. Westrom, L., Joesoef, R., Reynolds, G., Hagdu, A., and Thompson, S. E. Pelvic inflammatory disease and fertility. A cohort study of 1,844 women with laparoscopically verified disease and 657 control women with normal laparoscopic results. *Sex. Transm. Dis.* **19:**185–192, 1992.
4. Andrews, W. W., Goldenberg, R. L., Mercer, B., et al. The Preterm Prediction Study: association of second-trimester genitourinary chlamydia infection with subsequent spontaneous preterm birth. *Am. J. Obstet. Gynecol.* **183:**662–668, 2000.
5. Mardh, P. A., Influence of infection with *Chlamydia trachomatis* on pregnancy outcome, infant health and life-long sequelae in infected offspring. *Best Pract. Res. Clin. Obstet. Gynaecol.* **16:**847–864, 2002.
6. Peipert, J. F., Clinical practice. Genital chlamydial infections. *N. Engl. J. Med.* **349:**2424–2430, 2003.
7. Stary, A. European guideline for the management of chlamydial infection. *Int. J. STD AIDS* **12(Suppl. 3)**:30–33, 2001.
8. Mangione-Smith, R., O'Leary, J., and McGlynn, E. A. Health and cost-benefits of chlamydia screening in young women. *Sex. Transm. Dis.* **26:**309–316, 1999.
9. Davies, H. D., and Wang, E. E. Periodic health examination, 1996 update: 2. Screening for chlamydial infections. Canadian Task Force on the Periodic Health Examination. CMAJ **154:**1631–1644, 1996.
10. Scholes, D., Stergachis, A., Heidrich, F. E., Andrilla, H., Holmes, K. K., and Stamm, W. E. Prevention of pelvic inflammatory disease by screening for cervical chlamydial infection. *N. Engl. J. Med.* **334:**1362–1366, 1996.
11. Honey, E., Augood, C., Templeton, A., et al. Cost effectiveness of screening for *Chlamydia trachomatis*: a review of published studies. *Sex. Transm. Infect.* **78:**406–412, 2002.
12. Burstein, G. R., Zenilman, J. M., Gaydos, C. A., et al. Predictors of repeat *Chlamydia trachomatis* infections diagnosed by DNA amplification testing among inner city females. *Sex. Transm. Infect.* **77:**26–32, 2001.
13. Hook, E. W., and Handsfield, H. H. Gonococcal Infections of the Adult, in *Sexually Transmitted Diseases,* 3rd ed. J. N. Wasserheit J. N., ed, McGraw-Hill, New York, pp. 451–466, 1999.
14. Janda, W. M., and Knapp, J. S. Neisseria and Moraxella catarrhalis, in *Manual of Clinical Microbiology*. *Volume 1,* 8th ed. R. H. Yolken RH, ed. ASM, Washington, DC. pp. 585–608, 2003.
15. Bignell, C. J., European guideline for the management of gonorrhoea. *Int. J. STD AIDS* **12(Suppl. 3):**27–29, 2001.
16. Mertz, K. J., Levine W. C., Mosure, D. J., Berman, S. M., Dorian, K. J., and Hadgu, A. Screening women for gonorrhea: demographic screening criteria for general clinical use. *Am. J. Public Health* **87:**1535–1538, 1997.
17. Sherrard, J., and Barlow D. Gonorrhoea in men: clinical and diagnostic aspects. *Genitourin. Med.* **72:**422–426, 1996.
18. Iwen, P. C., Walker, R. A., Warren, K. L., Kelly, D. M., Hinrichs, S. H., and Linder, J. Evaluation of nucleic acid-based test (PACE 2C) for simultaneous detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in endocervical specimens. *J. Clin. Microbiol.* **33:**2587–2591, 1995.
19. Black, C. M., and Morse, S. A. The use of molecular techniques for the diagnosis and epidemiologic study of sexually transmitted infections. *Curr. Infect. Dis. Rep.* **2:**31–43, 2000.

20. Watson, E. J., Templeton, A., Russell, I., et al. The accuracy and efficacy of screening tests for *Chlamydia trachomatis:* a systematic review. *J. Med. Microbiol.* **51:**1021–1031, 2002.
21. Hadgu, A. The discrepancy in discrepant analysis. *Lancet* **348:**592–593, 1996.
22. Hadgu, A. Bias in the evaluation of DNA-amplification tests for detecting *Chlamydia trachomatis*. *Statist. Med.* **16:**1391–1399, 1997.
23. Miller, W. C. Bias in discrepant analysis: when two wrongs don't make a right. *J. Clin. Epidemiol.* **51:**219–231, 1998.
24. McAdam, A. J. Discrepant analysis: how can we test a test? *J. Clin. Microbiol.* **38:**2027–2029, 2000.
25. Black, C. M., Marrazzo, J., Johnson, R. E., et al. Head-to-head multicenter comparison of DNA probe and nucleic acid amplification tests for *Chlamydia trachomatis* infection in women performed with an improved reference standard. *J. Clin. Microbiol.* **40:**3757–3763, 2002.
26. Moncada, J., Schachter, J., Hook, E. W., 3rd, et al. The effect of urine testing in evaluations of the sensitivity of the Gen-Probe Aptima Combo 2 assay on endocervical swabs for *Chlamydia trachomatis* and *Neisseria gonorrhoeae:* the infected patient standard reduces sensitivity of single site evaluation. *Sex. Transm. Dis.* **31:**273–277, 2004.
27. Morris, R. E., Legault, J., and Baker, C. Prevalence of isolated urethral asymptomatic *Chlamydia trachomatis* infection in the absence of cervical infection in incarcerated adolescent girls. *Sex. Transm. Dis.* **20:**198–200, 1993.
28. Knapp, J. S., and Koumans, E. H. Neisseria and Branhamella, in *Manual of Clinical Microbiology,* 7th ed., R. H. Yolken ed. ASM, Washington, DC. pp. 586–603, 1999.
29. Van Der Pol, B., Quinn, T. C., Gaydos, C. A., et al. Multicenter evaluation of the AMPLICOR and automated COBAS AMPLICOR CT/NG tests for detection of *Chlamydia trachomatis*. *J. Clin. Microbiol.* **38:**1105–1112, 2000.
30. Van Der Pol, B., Ferrero, D. V., Buck-Barrington, L., et al. Multicenter evaluation of the BDProbeTec ET System for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in urine specimens, female endocervical swabs, and male urethral swabs. *J. Clin. Microbiol.* **39:**1008–1016, 2001.
31. Gaydos, C. A., Quinn, T. C., Willis, D., et al. Performance of the APTIMA Combo 2 assay for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in female urine and endocervical swab specimens. *J. Clin. Microbiol.* **41:**304–309, 2003.
32. Garrow, S. C., Smith, D. W., and Harnett, G. B. The diagnosis of chlamydia, gonorrhoea, and trichomonas infections by self obtained low vaginal swabs, in remote northern Australian clinical practice. *Sex. Transm. Infect.* **78:**278–281, 2002.
33. Cosentino, L. A., Landers, D. V., and Hillier, S.L. Detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* by strand displacement amplification and relevance of the amplification control for use with vaginal swab specimens. *J. Clin. Microbiol.* **41:**3592–3596, 2003.
34. Schachter, J., McCormack, W. M., Chernesky, M.A., et al. Vaginal swabs are appropriate specimens for diagnosis of genital tract infection with *Chlamydia trachomatis*. *J. Clin. Microbiol.* **41:**3784–3789, 2003.
35. Shafer, M. A., Moncada, J., Boyer, C. B., Betsinger, K., Flinn, S. D., and Schachter, J. Comparing first-void urine specimens, self-collected vaginal swabs, and endocervical specimens to detect *Chlamydia trachomatis* and *Neisseria gonorrhoeae* by a nucleic acid amplification test. J. Clin. Microbiol. **41:**4395–4399, 2003.
36. Serlin, M., Shafer, M. A., Tebb, K., et al. What sexually transmitted disease screening method does the adolescent prefer? Adolescents' attitudes toward first-void urine, self-collected vaginal swab, and pelvic examination. *Arch. Pediatr. Adolesc. Med.* **156:**588–591, 2002.
37. Richardson, E., Sellors, J. W., Mackinnon, S., et al. Prevalence of *Chlamydia trachomatis* infections and specimen collection preference among women, using self-collected vaginal swabs in community settings. *Sex. Transm. Dis.* **30:**880–885, 2003.
38. Oakeshott, P., Hay, P., Hay, S., et al. Detection of *Chlamydia trachomatis* infection in early pregnancy using self-administered vaginal swabs and first pass urines: a cross-sectional community-based survey. *Br. J. Gen. Pract.* **52:**830–832, 2002.
39. Wiesenfeld , H. C., Lowry, D. L., Heine, R. P., et al. Self-collection of vaginal swabs for the detection of Chlamydia, gonorrhea, and trichomoniasis: opportunity to encourage sexually transmitted disease testing among adolescents. *Sex. Transm. Dis.* **28:**321–325, 2001.
40. Hsieh, Y. H., Howell, M. R., Gaydos, J. C., McKee, K. T., Jr., Quinn, T. C., and Gaydos, C. A. Preference among female Army recruits for use of self-administrated vaginal swabs or urine to screen for *Chlamydia trachomatis* genital infections. *Sex. Transm. Dis.* **30:**769–773, 2003.
41. Golden, M. R., Astete, S. G., Galvan, R., et al. Pilot study of COBAS PCR and ligase chain reaction for detection of rectal infections due to *Chlamydia trachomatis*. *J. Clin. Microbiol.* **41:** 2174–2175, 2003.
42. Lister, N. A., Tabrizi, S. N., Fairley, C. K., and Garland, S. Validation of roche COBAS Amplicor assay for detection of *Chlamydia trachomatis* in rectal and pharyngeal specimens by an omp1 PCR assay. *J. Clin. Microbiol.* **42:**239–241, 2004.
43. Manavi, K., McMillan, A., and Young, H. The prevalence of rectal chlamydial infection amongst men who have sex with men attending the genitourinary medicine clinic in Edinburgh. *Int. J. STD AIDS* **15:**162–164, 2004.
44. Kessler, H. H., Pierer, K., Stuenzner, D., Auer-Grumbach, P., Haller, E. M., and Marth, E. Rapid detection of *Chlamydia trachomatis* in conjunctival, pharyngeal, and urethral specimens with a new polymerase chain reaction assay. *Sex. Transm. Dis.* **21:**191–195, 1994.
45. Ostergaard, L., Agner, T., Krarup, E, Johansen, U. B., Weismann, K., and Gutschik, E. PCR for detection of *Chlamydia trachomatis* in endocervical, urethral, rectal, and pharyngeal swab samples obtained from patients attending an STD clinic. *Genitourin. Med.* **73:**493–497, 1997.
46. Young, H., Anderson, J., Moyes A., and McMillan, A. Non-cultural detection of rectal and pharyngeal gonorrhoea by the Gen-Probe PACE 2 assay. *Genitourin. Med.* **73:**59–62, 1997.
47. Young, H., Manavi, K., and McMillan, A. Evaluation of ligase chain reaction for the non-cultural detection of rectal and pharyngeal gonorrhoea in men who have sex with men. *Sex. Transm. Infect.* **79:**484–486, 2003.
48. Davies, P. O., and Ridgway, G. L. The role of polymerase chain reaction and ligase chain reaction for the detection of *Chlamydia trachomatis*. *Int. J. STD.* AIDS **8:**731–738, 1997.
49. Ching, S., Lee, H., Hook, E. W., 3rd, Jacobs, M. R., and Zenilman, J. Ligase chain reaction for detection of *Neisseria gonorrhoeae* in urogenital swabs. *J. Clin. Microbiol.* **33:**3111–3114, 1995.
50. Carroll, K. C., Aldeen, W. E., Morrison, M., Anderson, R., Lee, D., and Mottice, S. Evaluation of the Abbott LCx ligase chain reaction assay for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in urine and genital swab specimens from a sexually transmitted disease clinic population. *J. Clin. Microbiol.* **36:**1630–1633, 1998.
51. Buimer, M., van Doornum, G. J., Ching, S., et al. Detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* by ligase chain reaction-based assays with clinical specimens from various sites: implications for diagnostic testing and screening. *J. Clin. Microbiol.* **34:**2395–2400, 1996.
52. Stary, A., Ching, S. F., Teodorowicz, L., and Lee, H. Comparison of ligase chain reaction and culture for detection of *Neisseria gonorrhoeae* in genital and extragenital specimens. *J. Clin. Microbiol.* **35:**239–242, 1997.
53. Gaydos, C. A., Theodore, M., Dalesio, N., Wood, B. J., and Quinn, T. C. Comparison of three nucleic acid amplification tests for detection of *Chlamydia trachomatis* in urine specimens. *J. Clin. Microbiol.* **42:**3041–3045, 2004.
54. Chan, E. L., Brandt, K., Olienus, K., Antonishyn, N., and Horsman, G. B. Performance characteristics of the Becton Dickinson ProbeTec System for direct detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in male and female urine specimens in comparison with the Roche Cobas System. *Arch. Pathol. Lab. Med.* **124:**1649–1652, 2000.
55. Palmer, H. M., Mallinson, H., Wood, R. L., and Herring, A. J. Evaluation of the specificities of five DNA amplification methods for the detection of *Neisseria gonorrhoeae*. *J. Clin. Microbiol.* **41:**835–837, 2003.

56. Culler, E. E., Caliendo, A. M., and Nolte, F. S. Reproducibility of positive test results in the BDProbeTec ET system for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae*. *J. Clin. Microbiol.* **41:**3911–3914, 2003.
57. Darwin, L. H., Cullen, A. P., Arthur, P. M., et al. Comparison of Digene hybrid capture 2 and conventional culture for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in cervical specimens. *J. Clin. Microbiol.* **40:**641–644, 2002.
58. Modarress, K. J., Cullen, A. P., Jaffurs, W. J., Sr., et al. Detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in swab specimens by the Hybrid Capture II and PACE 2 nucleic acid probe tests. *Sex. Transm. Dis.* **26:**303–308, 1999.
59. Schachter, J., Hook, E.W., 3rd, McCormack, W. M., et al. Ability of the Digene Hybrid Capture II test to identify *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in cervical specimens. *J. Clin. Microbiol.* **37:**3668–3671, 1999.
60. Darwin, L. H., Cullen, A. P., Crowe, S. R., Modarress, K. J., Willis, D. E., and Payne, W. J. Evaluation of the Hybrid Capture 2 CT/GC DNA tests and the GenProbe PACE 2 tests from the same male urethral swab specimens. *Sex. Transm. Dis.* **29:**576–580, 2002.
61. Van Der Pol, B., Williams, J. A., Smith, N. J., et al. Evaluation of the Digene Hybrid Capture II Assay with the Rapid Capture System for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae*. *J. Clin. Microbiol.* **40:**3558–3564.
62. Blanding, J., Hirsch, L., Stranton, N., et al. Comparison of the Clearview Chlamydia, the PACE 2 assay, and culture for detection of *Chlamydia trachomatis* from cervical specimens in a low-prevalence population. *J. Clin. Microbiol.* **31:**1622–1625, 1993.
63. Clarke, L. M., Sierra, M. F., Daidone, B. J., Lopez, N., Covino, J. M., and McCormack, W. M. Comparison of the Syva MicroTrak enzyme immunoassay and Gen-Probe PACE 2 with cell culture for diagnosis of cervical *Chlamydia trachomatis* infection in a high-prevalence female population. *J. Clin. Microbiol.* **31:**968–971, 1993.
64. Iwen, P. C., Blair, T. M., and Woods, G. L. Comparison of the Gen-Probe PACE 2 system, direct fluorescent-antibody, and cell culture for detecting *Chlamydia trachomatis* in cervical specimens. *Am. J. Clin. Pathol.* **95:**578–582, 1991.
65. Limberger , R. J., Biega, R., Evancoe, A., McCarthy, L., Slivienski, L., and Kirkwood, M. Evaluation of culture and the Gen-Probe PACE 2 assay for detection of *Neisseria gonorrhoeae* and *Chlamydia trachomatis* in endocervical specimens transported to a state health laboratory. *J. Clin. Microbiol.* **30:**1162–1166, 1992.
66. Stary, A., Kopp, W., Zahel, B., Nerad, S., Teodorowicz, L., and Horting-Muller, I. Comparison of DNA-probe test and culture for the detection of *Neisseria gonorrhoeae* in genital samples. *Sex. Transm. Dis.* **20:**243–247, 1993.
67. Vlaspolder, F., Mutsaers, J. A., Blog, F., and Notowicz, A. Value of a DNA probe assay (Gen-Probe) compared with that of culture for diagnosis of gonococcal infection. *J. Clin. Microbiol.* **31:**107–110, 1993.
68. Hale, Y. M., Melton, M. E., Lewis, J. S., and Willis , D. E. Evaluation of the PACE 2 *Neisseria gonorrhoeae* assay by three public health laboratories. *J. Clin. Microbiol.* **31:**451–453, 1993.
69. Lewis, J. S., Fakile, O., Foss, E., et al. Direct DNA probe assay for *Neisseria gonorrhoeae* in pharyngeal and rectal specimens. *J. Clin. Microbiol.* **31:**2783–2785, 1993.
70. Boyadzhyan, B., Yashina, T., Yatabe, J. H., Patnaik, M., and Hill, C.S. Comparison of the APTIMA CT and GC assays with the APTIMA combo 2 assay, the Abbott LCx assay, and direct fluorescent-antibody and culture assays for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae*. *J. Clin. Microbiol.* **42:**3089–3093, 2004.
71. Martin, D. H., Cammarata, C., Van Der Pol, B., et al. Multicenter evaluation of AMPLICOR and automated COBAS AMPLICOR CT/NG tests for *Neisseria gonorrhoeae*. *J. Clin. Microbiol.* **38:**3544–3549, 2000.
72. Hammerschlag, M. R., Roblin P. M., Gelling, M., Tsumura, N., Jule, J. E., and Kutlin, A. Use of polymerase chain reaction for the detection of *Chlamydia trachomatis* in ocular and nasopharyngeal specimens from infants with conjunctivitis. *Pediatr. Infect. Dis. J.* **16:**293–297, 1997.
73. Koumans, E. H., Black, C. M., Markowitz, L. E., et al. Comparison of methods for detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* using commercially available nucleic acid amplification tests and a liquid pap smear medium. *J. Clin. Microbiol.* **41:**1507–1511, 2003.

35 Human Papillomavirus

THERESA M. VOYTEK AND GREGORY J. TSONGALIS

1. INTRODUCTION

Once the most common cause of cancer death in women in the United States, the incidence and mortality rates from cervical cancer declined dramatically with the introduction of the Papanicolaou (Pap) smear *(1)*. Absence of such screening programs in many developing countries, however, is the most significant factor responsible for the high incidence and mortality rates of this cancer. Excluding the common skin cancers, cervical cancer remains the second most common cancer in women worldwide *(2)*.

Risk factors for cervical cancer include young age at first intercourse, number of sexual partners, parity, smoking, oral contraceptive use, and presence of high-risk human papillomavirus (HPV) type *(3)*. Of these risk factors, the type of HPV and persistence of infection have become recognized as the most important *(4)*. Recently, it has been shown that more than 99% of all cervical cancers have detectable HPV DNA sequences *(5)*.

2. HUMAN PAPILLOMAVIRUS

Human papillomavirus is a double-stranded tumor virus that belongs to the larger papovavirus family. It is recognized as a ubiquitous virus that has been detected in epithelial lesions from a variety of animals. There are more than 100 types of HPV that are known to infect cutaneous or mucosal sites in humans; some viral types only infect oral, whereas others infect both oral and genital sites. HPV is characterized as a small, icosahedral, nonenveloped DNA virus whose genome consists of approx 8000 basepairs (bp) and codes for 10 viral proteins, 8 early gene products, and 2 late gene products *(6)*. The early gene products have been shown to be involved with viral replication and tumorigenesis, whereas the late gene products code for structural proteins of the viral capsid.

Early gene proteins E6 and E7 are oncoproteins with transforming potential. These two viral proteins have been shown to bind to and inactivate the cellular p53 and the Rb tumor suppressor gene products *(5)* (Fig. 1). p53 degradation is induced upon binding of the E6 protein, whereas E7 alters the phosphorylation of the Rb gene product leading to its inactivation *(7,8)*. E2, another of the early HPV gene proteins, encodes a regulatory protein that affects expression of E6 and E7. During integration of HPV, the E2 gene is disrupted and leads to overexpression of E6 and E7 *(9)*.

The majority of HPV types are classified as nongenital and are mainly associated with benign cutaneous lesions such as hand and plantar warts. Of the 25–30 genital HPV types, the more common low-risk types include types 6, 11, 42, 43, and 44 and are associated with low-grade lesions. The high-risk types (HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68) are associated with low-grade and high-grade cervical lesions and cancer. Because more than 99% of all cervical cancers contain high-risk HPV DNA, the analysis of cervical specimens for the presence of HPV has become a standard screening protocol.

3. CERVICAL CANCER AND THE PAP TEST

Human papillomavirus is the major cause of carcinoma and precancerous lesions of the cervix *(10)*. Although cervical cancer was once the leading cause of cancer related death in women, it now ranks 13 *(11)*, mostly because of the widespread use of the Pap test. The lesions that lead to cervical cancer (Fig. 2) are usually detectable using the Pap test and are highly curable; however, the Pap test is not a perfect screening test. The sensitivity of a single conventional Pap test has been estimated to be 51% *(12)*. False-negative Pap tests have many causes, including inadequate sampling, obscuring blood, inflammation, or other factors obscuring adequate cell visualization, only few cells on the slide, and screening and interpretive errors. The liquid-based Pap test, a method in which the sample is placed directly into a vial of fixative rather than smeared directly onto a glass slide, has improved the quality of the Pap test by providing a cleaner, better preserved preparation than the conventional Pap smear and has improved the sensitivity of the Pap test *(13–17)*. However, although the liquid-based test is more sensitive in detecting cervical abnormalities than the conventional Pap smear, the Pap test alone is still unable to detect all significant cervical lesions. Furthermore, many Paps have equivocal results reported as "atypical squamous cells of undetermined significance" (ASC-US) by The Bethesda System for reporting cervical/vaginal cytologic diagnoses *(18)*. Although some women with ASC-US diagnoses have significant cervical preneoplastic lesions, many with only benign conditions undergo

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

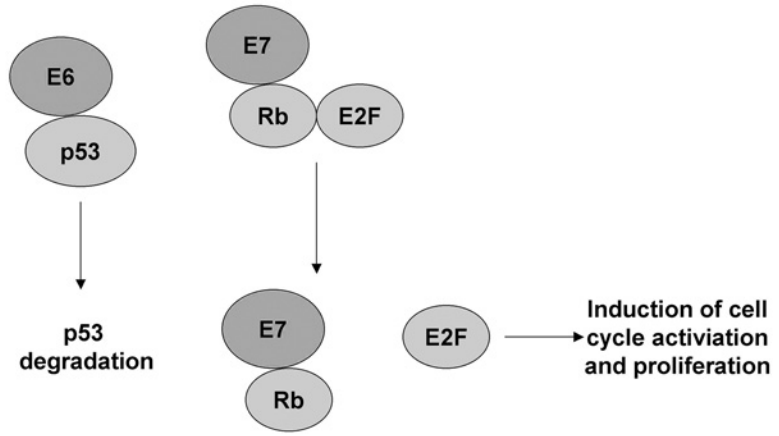


Fig. 1. Mechanism of HPV-induced oncogenesis. (Figure appears in color in accompanying CD ROM.)

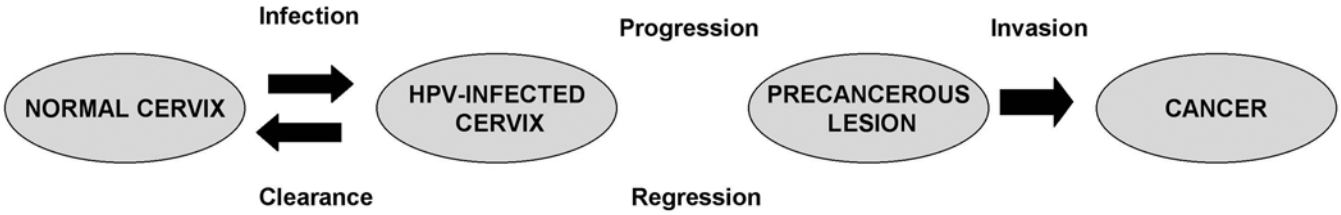


Fig. 2. Cervical carcinogenesis. (Figure appears in color in accompanying CD ROM.)

cervical colposcopy and biopsy to follow up the abnormal Pap test, resulting in unnecessary cost and patient anxiety.

4. HPV AND CERVICAL CANCER

Several recent studies have evaluated the utility of HPV testing as a possible adjunct test to increase the accuracy of the Pap test. HPV is a sexually transmitted virus that is most prevalent in young women of childbearing age. There are over 100 types that are categorized as either high-risk, oncogenic types, or low-risk types based on their ability to cause high-grade squamous intraepithelial lesion (HSIL) *(19)*. Approximately 13 types of HPV are associated with cervical carcinoma, including types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68. Most HPV infections are transient *(20,21)*. A study of HPV infections in young women aged 13 to 22 showed that 70% of high-risk HPV infections and over 90% of low-risk infections regress within 3 yr *(20)*. A small percentage of high-risk HPV infections are persistent and progress to HSIL or carcinoma. The progression rate is usually slow, requiring several years to progress from a low-grade to a high-grade lesion.

5. THE ASCUS LOW GRADE TRIAGE STUDY

Recent studies have shown HPV testing to be very useful in the management of equivocal cervical cytologic diagnoses *(22)*. The largest of these is the ASCUS Low Grade Triage Study (ALTS) *(23–25)*. This study was initiated by the National Cancer Institute because of the lack of consensus among physicians regarding the optimal management approach for ASC-US and low-grade squamous intraepithelial lesion (LSIL). Although there was agreement among health care providers that Pap tests showing high-grade lesions require colposcopic examination, the further evaluation of ASC-US and LSIL was variable with some physicians following with repeat Pap tests and others with colposcopy and cervical biopsy. The result was that many patients without clinically significant lesions underwent cervical biopsy at substantial cost and undue anxiety. ALTS was conducted to try to establish a cost-effective management approach that would be sensitive in detecting significant cervical lesions while limiting the rate of false-positive tests. In this large multicenter trial, 3488 patients with ASC-US and 1572 patients with LSIL were randomly assigned to 1 of 3 research arms between November 1996 and December 1998. The three management strategies were (1) immediate colposcopy, (2) HPV DNA testing by the Hybrid Capture 2 (HC2) method (Digene Corporation, Gaithersburg, MD), and (3) follow-up with cytology alone. Early in the trial, the HPV triage arm for women referred for LSIL was closed because of a very high prevalence of high-risk HPV in this population, preventing the effective triage of these patients *(25)* In other words, so many patients with LSIL on the Pap test were positive for high-risk HPV DNA that it was not useful in triaging patients for colposcopy. Consequently, the current recommendation for women with a cytologic diagnosis of LSIL is colposcopic examination and cervical biopsy of any clinical abnormality to evaluate for a high-grade lesion.

Human papillomavirus DNA testing proved very useful in triaging patients with ASC-US on cytology. Testing for high-risk HPV DNA referred 56% of patients for colposcopy and identified 96% of women with severe squamous dysplasia, cervical intraepithelial lesion grade 3 (CIN3). Repeat Pap test using a threshold of HSIL referred only 8% of patients for colposcopy but identified only 44% of patients with CIN3. Using a lower cytology threshold of ASC-US for colposcopy referral increased the Pap detection rate of CIN3 to 85%, referring 58% of women

to colposcopy. This study showed HPV testing to be an effective method of evaluation of patients with ASC-US. The advantages of this approach compared to repeat Pap include (1) the ability to test for HPV on the original sample, avoiding a second office visit to obtain a second specimen, (2) decreasing patient anxiety for those patients with a negative HPV result, and (3) a very sensitive method of identifying women at risk of harboring a clinically significant cervical lesion.

The ALTS also demonstrated the poor reproducibility of cytologic diagnoses *(26)*. In the review of cytology, there were significant discrepancies in interpretations among pathologists both on the Pap tests and the tissue biopsies. Pap diagnoses of ASC-US and LSIL in particular showed poor reproducibility. These findings also supported the use of HPV DNA testing to provide additional information in equivocal cases.

6. USES OF HPV TESTING

The utility of HPV testing as a reflex test in cases of ASC-US is now widely accepted and the indications for HPV testing are expanding further. HPV DNA testing was approved by the Food and Drug Administration (FDA) as a primary screening test, in conjunction with the Pap test, for women over the age of 30. This combined test, the DNAwithPAP test (Digene Corporation) provides almost 100% sensitivity and negative predictive value. It is more accurate than either test alone and is cost-effective because it allows less frequent screening without compromising patient care *(27)*. The American Cancer Society recently published guidelines that include this approach *(28)*. HPV DNA testing as a primary screening test is generally not recommended in women less than 30 yr of age because of the high prevalence of transient HPV infection in this population. Screening of this group would result in an excess of patients receiving colposcopy for low-grade lesions that would spontaneously regress without treatment. Other reported indications for HPV DNA testing are (1) to follow patients with HPV-positive ASCUS, (2) to follow those with LSIL on Pap after colposcopy is negative for a high-grade lesion, and (3) to follow patients after treatment of a high-grade lesion to provide more evidence of cure *(23,29–31)*. In summary, the potential uses of HPV DNA testing are as follows:

1. Triage of patients with ASCUS on Pap test
2. Primary screening combined with Pap test every 3 yr in women over the age of 30
3. Management of women with LSIL or HPV DNA positive ASC-US on Pap and negative colposcopy for a high-grade lesion
4. Follow-up of patients after treatment of high-grade lesion to ensure absence of residual infection
5. Additional work-up of cases with discrepant cytology, colposcopy, and biopsy findings, to assist in making appropriate management decisions

7. DETECTION OF HPV

Human papillomavirus and its related tissue changes can be detected visually at clinical colposcopic examination, morphologically by histopathologic examination or by DNA analysis. Viral detection relies on these other methods, as the virus itself does not grow in culture. Using traditional molecular diagnostic approaches, testing for HPV can be performed utilizing probe hybridization techniques, signal amplification, or target amplification methods *(32)*.

Table 1
HPV Genotypes Detected by the Digene HC2 Assay

Low-risk HPV types	6, 11, 42, 43, 44
High-risk HPV types	16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68

8. DIRECT PROBE HYBRIDIZATION

Initial testing for HPV was performed using the Southern blot transfer technology and radioisotopic labeled probe *(33)*. This assay required extraction of nucleic acids from the specimen, restriction enzyme digestion, electrophoresis, transfer, hybridization, and signal detection. Another approach to using direct probes was in combination with a second blotting technique known as a dot or slot blot. This technique does not require enzyme digestion or electrophoresis and can evaluate numerous specimens by applying the nucleic acids directly to a membrane with subsequent hybridization with a probe of interest. Finally, probes have also been applied directly to tissue sections or cells on a slide as an *in situ* hybridization (ISH) assay *(34–36)*. The Ventana Inform HPV assay is a chromogenic ISH test that can differentiate between episomal or cytoplasmic viral replication and viral integration *(37)*. Overall, blotting approaches have been disadvantageous with respect to labor intensiveness, turnaround times, and throughput. In addition, quantities of DNA needed for some of these technologies, specimen types, and low sensitivities do not favor the widespread use of these testing modalities. With improvements in chemical detection systems and automation, such technologies might become more widely used.

9. DIGENE HC2

Signal amplification technologies are dependent on the hybridization of probe sequences to the target nucleic acid sequences that are then detected by multiple reporter molecules. Because numerous reporter molecules can attach to each probe:target hybrid, the sensitivity of this assay is greatly increased compared to a direct probe assay. Currently, the only FDA-cleared assay for HPV analysis is the Hybrid Capture 2 (HC2) assay (Digene Corporation, Inc, Gaithersburg, MD). HC2 can be used for analysis of cervical brush, fresh/frozen cervical biopsy, and liquid cytology specimens. This assay is a two-tiered test with a cocktail of 13 probes for the high-risk oncogenic HPV types and a cocktail of 5 probes for the low-risk nononcogenic HPV types (Table 1). This test utilizes a chemiluminescent-based detection system for identifying RNA probes that have hybridized to individual HPV DNA target sequences *(38)*. This RNA : DNA hybrid construct is detected by proprietary antibodies that are conjugated to reporter molecules. The steps involved in this assay are shown in Fig. 3.

Multiple alkaline phosphatase-conjugated antibodies bind to the RNA : DNA hybrid and result in an amplified chemiluminescent signal. This signal is measured in relative light units

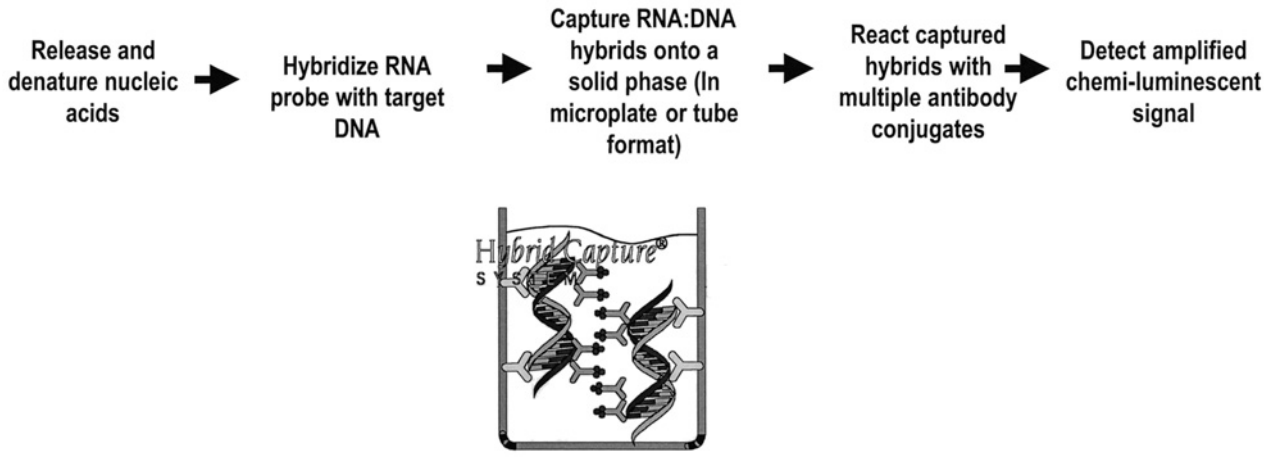


Fig. 3. Hybrid Capture 2 technology.

(RLUs). All specimen RLU values are converted into a ratio with the appropriate positive cutoff value (RLU/CO). If the RLU/CO is >1, then a positive result is obtained for that specimen. This result indicates the presence of one of the HPV types detected by the cocktail of probes used. It does not identify the specific type of HPV. There has been some controversy as to the absolute cutoff of 1 RLU. Most laboratories have established equivocal ranges so that specimens whose RLU/CO ratio fall within that range are repeated. These equivocal ranges usually result in repeating specimens with RLU/CO between 1.0 and 3.0.

10. POLYMERASE CHAIN REACTION

Assays utilizing target amplification, on the other hand, are based on the *in vitro* amplification of a target nucleic acid sequence. Polymerase chain reaction (PCR) represents the most commonly used of these technologies. In these assays, a target sequence is identified by small oligonucleotides known as primers. Synthesis of multiple copies of this target sequence is accomplished by a polymerase enzyme during optimal temperature cycling. The amplified target sequence is then detected using a variety of methods, including direct sequencing and probe-based assays. Many PCR assays for HPV detection have used type-specific or consensus primers *(39)*. The type-specific assays typically identify sequence variations in the E6 or E7 genes, whereas the consensus primers target conserved regions in the L1 capsid gene. There are many advantages to the PCR technology for such screening applications, including automation capabilities, turnaround time, multiplexing, sensitivity/specificity, multiple specimen types, and small-specimen volume. The major concern with the in vitro amplification technologies is the potential for contamination. Much of this has been overcome by either enzymatic (uracil-*N*-glycosylase) or physical (closed system) contamination control.

Performance of these assays varies and is dependent on traditional diagnostic testing variables. Critical to diagnostic testing are the analytical sensitivity and specificity of the assay, which determine the lowest detectable quantity of DNA and the ability of the assay to detect the intended target, respectively. Amplification methods, both target and signal, have much higher sensitivities than direct probe methods.

11. CONCLUSION

Although the HC2 assay is currently the only FDA-approved assay for HPV screening, technologies continue to evolve for many molecular diagnostic applications. Other amplification and detection technologies, including macroarrays as well as liquid- and microarray-based assays, will be developed for such high-throughput screening applications. In addition to these technological advances, the role of HPV screening in patient management continues to evolve with respect to indications for testing, need for type-specific results, and clinical utility of viral load testing.

REFERENCES

1. Greenlee, R. T., Hill-Harmon, M. B., Murray, T., and Thun, M. Cancer statistics 2001. *CA Cancer J. Clin.* **50:**7–33, 2001.
2. Boffeta, P. and Parkin, D. M. Cancer in developing countries. *CA Cancer J. Clin.* **44:**81–90, 1994.
3. Adam, E., Berkova, Z., Daxnerova, Z., Icenogle, J., Reeves, W. C., and Kaufman, R. H. Papillomavirus detection: demographic and behavioral characterisitics influencing the identification of cervical disease. *Am. J. Obstet. Gynecol.* **182:**257–264, 2000.
4. Wallin, K. L., Wiklund, F., Angstrom, T., et al. Type-specific persistence of human papillomavirus DNA before the development of invasive cervical cancer. *N. Engl. J. Med.* **341:**1633–1638, 1999.
5. Kaufman, R. H., Adam, E., and Vonka, V. Human papillomavirus infection and cervical carcinoma. *Clin. Obstet. Gynecol.* **43:** 363–380, 2000.
6. Stoler, M. H. Human papillomavirus biology and cervical neoplasia. *Arch. Pathol. Lab. Med.* **127:**935-939, 2003.
7. Scheffner, M., Huibregtse, J. M., Vierstra, R. D., et al. The HPV16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. *Cell* **75:**495–505, 1993.
8. Chellappan, S., Kraus, V. B., Kroger, B., et al. Adenovirus E1A, simian virus 40 tumor antigen, and human papillomavirus E7 protein share the capacity to disrupt the interaction between the transcription factor E2F and the retinoblastoma gene product. *Proc. Natl. Acad. Sci. USA.* **89:**4549–4553, 1992.
9. Choo, K. B., Pan, C. C., and Han, S. H. Integration of human papillomavirus type 16 into cellular DNA of cervical carcinoma: preferential deletion of the E2 gene and invariable retention of the long control region and the E6/E7 open reading frames. *Virology* **161:**259–261, 1987.
10. Munoz, N., Bosch, F. X., de Sanjose, S., et al. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N. Engl. J. Med.* **348:**518–527, 2003.
11. Jemal, A., Thomas, A., Murray, T., and Thun, M. Cancer statistics, 2002. *CA Cancer J. Clin.* **52:**23–47, 2002.

12. Bastian, L., Datta, S., Hasselblad, V., Hickey, J., Myers, E., and Nanda, K. *Evaluation of Cervical Cytology.* Evidence Report/ Technology Assessment Number 5. Department. of Health and Human Services, Agency for Health Care Policy and Research, Bethesda, MD, 1999.
13. Papillo, J. L., Zarka, M. A., and St. John, T. L. Evaluation of the ThinPrep Pap Test in clinical practice. A seven-month, 16,314-case experience in northern Vermont. *Acta Cytol.* **42:**203–208, 1998.
14. Diaz-Rosario, L. A. and Kabawat, S. E. Performance of a fluid-based, thin layer Papanicolaou smear method in the clinical setting of an independent laboratory and an outpatient screening population in New England. *Arch. Pathol. Lab. Med.* **123:**817–821, 1999.
15. Weintraub, J. and Morabia, A. Efficacy of a liquid-based thin layer method for cervical cancer screening in a population with a low incidence of cervical cancer. *Diagn. Cytopathol.* **22:**52–59, 2000.
16. Linder, J. and Zahniser, D. ThinPrep Pap testing to reduce false-negative cytology. *Arch. Pathol. Lab. Med.* **122:**139–144, 1998.
17. Hutchinson, M. L., Zahniser, D. J., Sherman, M. E., et al. Utility of liquid-based cytology for cervical carcinoma screening: results of a population-based study conducted in a region of Costa Rica with a high incidence of cervical carcinoma. *Cancer* **87:**48–55, 1999.
18. Solomon, D., Davey, D., Kurman, R., et al. The 2001 Bethesda System: terminology for reporting results of cervical cytology. *JAMA* **287:**2114–2119, 2002.
19. Schiffmann, M. and Castle, P. E. Human papillomavirus: epidemiology and public health. *Arch. Pathol. Lab. Med.* **127:**930–934, 2003.
20. Moscicki, A. B., Shiboski, S., Broering, J., et al. The natural history of human papillomavirus infection as measured by repeated DNA testing in adolescent and young women. *J. Pediatr.* **132:**277–284, 1998.
21. Ho, G. Y., Bierman, R., Beardsley, L., et al. Natural history of cervicovaginal papillomavirus infection in young women. *N. Engl. J. Med.* **338:**423–428, 1998.
22. Manos, M. M., Kinney, W. K., Hurley, L. B., et al. Identifying women with cervical neoplasia: using human papillomavirus DNA testing for equivocal Papanicolaou results. *JAMA* **281:**1605–1610, 1999.
23. Shiffman, M. and Solomon, D. Findings to date from the ASCUS-LSIL triage study (ALTS). *Arch. Pathol. Lab. Med.* **127:**946–949, 2003.
24. Solomon, D., Schiffman, M., Tarone, R., et al. Comparison of three management strategies for patients with atypical squamous cells of undetermined significance: baseline results from a randomized trial. *J. Natl. Cancer Inst.* **93:**293–299, 2001.
25. The ALTS Group. HPV testing as a triage strategy for women with cytologic evidence of low grade squamous intraepithelial lesions: early results from a randomized trial on the clinical management of ASCUS and LSIL cytologic abnormalities. *J. Natl. Cancer Inst.* **92:**397–402, 2000.
26. Stoler, M. H., Schiffman, M. for the ALTS Group. Interobserver reproducibility of cervical cytologic and histologic diagnoses: realistic estimates from the ASCUS-LSIL Triage Study (ALTS). *JAMA* **285:**1500–1505, 2001.
27. Richart, R. M. and Lorincz, A. T. Human papillomavirus DNA testing as an adjunct to cytology in cervical screening programs. *Arch. Pathol. Lab. Med.* **127:**959–968, 2003.
28. Saslow, D., Runowicz, C. D., Solomon, D., et al. American Cancer Society guideline for the early detection of cervical neoplasia and cancer. *CA Cancer J. Clin.* **52:**342–362, 2002.
29. Wright, T. C., Cox, J. T., and Massad, L. S., for the ASCCP-sponsored Consensus Conference. 2001 consensus guidelines for the management of women with cervical cytological abnormalities. *JAMA* **287:**2120–2126, 2002.
30. Cox, J. T. Clinical role of HPV testing. *Obstet. Gynecol. Clin. North Am.* **23:**811–851, 1996.
31. Paraskevaidis, E., Koliopoulos, F., Alamanos, Y., Malamou-Mitsi, V., Lolis, C. D., and Kitchener, H. C. Human papillomavirus testing and the outcome of treatment for cervical intraepithelial neoplasia. *Obstet. Gynecol.* **98:**833–836, 2001.
32. Hubbard, R. A. Human papillomavirus testing methods. *Arch. Pathol. Lab. Med.* **127:**940–945, 2003.
33. Schiffman, M. N., Bauer, H. M., Lorincz, A., et al. Comparison of Southern blot hybridization and polymerase chain reaction methods for the detection of human papillomavirus DNA. *J. Clin. Microbiol.* **29:**573–577, 1991.
34. Caussy, D., Orr, W., Daya, A. D., Roth, P., Reeves, W., and Rawls, W. Evaluation of methods for detecting human papillomavirus deoxyribonucleotide sequences in clinical specimens. *J. Clin. Microbiol.* **26:**236–243, 1988.
35. Bekkers, R. L., Melchers, W. J., Bulten, J., et al. Localized distribution of human papillomavirus genotypes in the uterine cervix. *Eur. J. Gynaecol. Oncol.* **23:**203–206, 2002.
36. Huang, C. C., Qiu, J. T., Kshima, M. L., Kurman, R. J., and Wu, T. C. Generation of type specific probes for the detection of single copy human papillomavirus by a novel in situ hybridization method. *Mod. Pathol.* **11:**971–977, 1998.
37. Bolick, D. R. Laboratory implementation of human papillomavirus testing. *Arch. Pathol. Lab. Med.* **127:**984–990, 2003.
38. Clavel, C., Massure, M., Bory, J. P., et al. Hybrid capture II-based human papillomavirus detection, a sensitive test to detect in routine high grade cervical lesions: a preliminary study on 1518 women. *Br. J. Cancer* **80:**1306–1311, 1999.
39. Husnjak, K., Grce, M., Magdic, L., and Paveliae, K. Comparison of five different polymerase chain reaction methods for the detection of human papillomavirus in cervical cell specimens. *J. Virol. Methods* **88:**125–134, 2000.

36 Molecular Diagnostics for HIV-1

ANGELA M. CALIENDO AND CHARLES E. HILL

1. INTRODUCTION

An estimated 30 million people are currently infected with the human immunodeficiency virus type 1 (HIV-1), and over the last two decades there have been more than 20 million deaths from the acquired immunodeficiency syndrome (AIDS). In addition, approx 16,000 new infections occur daily, making this a worldwide epidemic. In 1983, Luc Montgnier and his colleagues at the Institute Pasteur and Robert Gallo and his colleagues at the US National Institutes of Health independently identified HIV-1 as the cause of AIDS *(1)*. By 1985, an antibody test was developed to identify infected individuals. Over the following year zidovudine (AZT), the first anti-HIV drug, went on clinical trial and received Food and Drug Administration (FDA) approval in record time. In 1986, a second type of HIV called HIV-2 was isolated from AIDS patients in West Africa *(2)*. Although additional drugs such as ddI and ddC were developed, the next major breakthrough did not occur until a decade later with the discovery of a new class of antiviral agents, the protease inhibitors; along with this came the implementation of combination therapy for treatment of HIV-infected individuals. Despite the great progress in diagnosis, treatment, and prevention, HIV remains a major cause of death in most parts of the world. Thus far, the pandemic has led to over 70 million infections and an annual worldwide mortality of 3 million people *(1)*. In addition, the emergence of HIV resistance has been described for all classes of antiretroviral agents *(3–7)*. Recent studies by Little and others *(8,9)* have shown that a significant number of newly infected HIV patients harbor viruses that are resistant to at least one drug and a small, yet significant, proportion of these patients harbor viruses that are resistant to all available antiviral agents.

2. VIROLOGY

Both HIV-1 and HIV-2 are members of the genus *Lentivirinae* of the family of Retroviridae. They are enveloped positive-stranded RNA viruses, and the virion contains two identical copies of the single-stranded RNA of approx 9–10 kb in length and the enzyme reverse transcriptase. The envelope contains virally encoded glycoproteins that mediate the adsorption and penetration into the host cell via the chemokine receptors CCR5 and CXCR4. Following penetration into the cell, reverse transcriptase converts the viral RNA into DNA, which enables HIV to be integrated into the host cell's DNA *(10)*. It is estimated that an infected cell might generate as many as 10^{10} viral particles every day. Considering the high rate of viral replication and the error-prone nature of the reverse transcriptase enzyme, a myriad of viral strains (quasispecies) circulate in a given host and in the community *(2,10)*. Based on nucleic acid sequences, HIV-1 isolates have been classified into two groups: group M (major) and group O (other) viruses. Within the M group, at least nine subtypes (A–H, J) have been identified, whereas there have been only five subtypes of HIV-2 (A–E) reported *(2,10)*.

3. HIV-1 VIRAL LOAD TESTING

3.1. CLINICAL UTILITY As antiretroviral therapies were being introduced and shown to be quite effective, it became apparent that there was a need to develop methods for quantifying the amount of virus in patient specimens. HIV-1 measurements are very useful for predicting time to progression to AIDS, determining when to initiate antiretroviral therapy, and monitoring response to therapy. The availability of commercial assays combined with their proven clinical utility made viral-load testing the standard of care in managing patients with HIV-1 infection. Plasma HIV-1 RNA levels are a strong predictor of a rapid progression to AIDS that is independent of CD4 cell count *(11)*. In one study of homosexual men with documented time of seroconversion, an initial viral load of 100,000 equivalents/mL soon after seroconversion was associated with a greater than 10-fold increase in the risk for developing AIDS. However, persons who maintained a viral load of <1000 equivalents/mL did not progress to AIDS during the next 5 yr *(11)*. Even in patients in which the duration of infection is unknown, the level of plasma RNA can be used to predict prognosis. A study of 180 seropositive men showed patients with a baseline RNA level of <4530 molecules/mL had a median time to progression to AIDS of >10 yr. In contrast, for persons with a baseline viral load of >36,270 molecules/mL, the time to progression to AIDS was 3.5 yr *(12)*.

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

Table 1
Molecular Assays for Quantification of HIV-1 RNA

Assay	*Method*	*Target*	*Range*[a]
Amplicor HIV-1 Monitor Test v1.5, Standard (Roche Diagnostics , Indianapolis, IN)	RT-PCR	HIV-1 *gag* gene	400–750,0000 copies/mL
Amplicor HIV-1 Monitor Test v 1.5, Ultrasensitive (Roche Diagnostics , Indianapolis, IN)	RT-PCR	HIV-1 *gag* gene	50–100,000 copies/mL
Versant HIV-1 RNA 3.0 Assay (bDNA) (Bayer Diagnostics Corporation, Tarrytown, NY)	Branched DNA	HIV-1 *pol* gene	75–500,000 copies/mL
NucliSens HIV-1 QT Assay (BioMérieux, Inc., Durham, NC)	NASBA	HIV-1 *gag* gene	176–3,470,000 copies/mL

Abbreviations: RT-PCR = reverse transcription–polymerase chain reaction; NASBA = nucleic acid sequence-based amplification.
[a]See text for a detailed discussion of the dynamic range of the assays.

In addition to prognostic utility, the measurement of HIV RNA levels is useful for monitoring response to antiretroviral therapy. The magnitude of the decrease in viral load in response to therapy is dependent on the specific antiretroviral therapy used. Decreases in viral load of approx 2 $\log_{10}$ have been reported with the currently used regimens, which include combinations of nucleoside analog inhibitors with either protease inhibitors or non-nucleoside reverse transcriptase inhibitors *(13–15)*. The ultimate goal of therapy is to achieve a viral load below the limit of detection of the assay, as studies have shown that the suppression of plasma viral load to below 20 copies/mL was associated with a longer response to antiretroviral therapy compared with the duration of response achieved when viral load was suppressed to below 500 copies/mL *(16)*. Guidelines for the use of HIV-1 RNA levels in clinical practice have been published [*(17)*; http://www.aidsinfo. nih.gov, http://www.iasusa.org]. In general, plasma HIV-1 RNA levels are measured immediately before beginning therapy and 2–8 wk after start of therapy to determine initial response. Testing is repeated at 3- to 4-mo intervals to evaluate continued effectiveness of the regimen, although more frequent testing is done if the patient has a deteriorating clinical course. Viral-load measurements are used in conjunction with CD4 cell counts to manage HIV-infected persons. Although CD4 cell counts are inversely correlated with viral load, the association is weak. For any given CD4 cell count, there might be a 3 $\log_{10}$ range in viral load. Even with this modest correlation, CD4 cell counts remain very useful in determining when to initiate both antiretroviral therapy and prophylactic therapy for opportunistic infections.

3.2. COMMERCIALLY AVAILABLE ASSAYS Three commercial HIV-1 viral-load assays are available for clinical use and are listed in Table 1. The Amplicor reverse transcription– polymerase chain reaction [RT-PCR] assay, the NucliSens nucleic acid sequence-based amplification [NASBA] assay, and the Versant branched DNA [bDNA] assay are all FDA approved. These assays have been modified to have a lower limit of quantification of approx 50–100 copies/mL; however, the manner of determining this lower limit varies. Some manufacturers set the lower limit of quantification at the amount of nucleic acid that can be reliably quantified in 50% of replicate samples, whereas others require that it be achieved in 95% of replicates. In an effort to standardize terminology, the limit of quantification is best determined based on 95% of replicates having quantifiable levels of nucleic acid. Although the lower limit of quantification can be determined by running a large number of specimens at or near the quantity that becomes undetectable, it can be calculated using probit analysis of data from significantly fewer replicates.

3.3. BAYER QUANTIPLEX® HIV RNA V 3.0 ASSAY The QUANTIPLEX® HIV RNA v 3.0 ASSAY (Bayer Nucleic Acid Diagnostics, Norwood, MA) is a signal amplification test that relies on a sandwich nucleic acid hybridization procedure for the quantification of HIV-1 RNA. Initially, the HIV-1 genomic RNA is released from the virions and captured onto the surface of a microwell by a set of immobilized HIV-specific oligonucleotide target probes. A second set of target probes are included that hybridize to both the viral RNA and to the preamplifier probes. The preamplifier probes hybridize to the bDNA amplifiers. Multiple copies of an alkaline phosphatase-labeled probe complementary to the amplifier DNA sequence are then hybridized to the immobilized complex to further amplify the signal. Detection is achieved by incubating the complex with a chemiluminescent substrate and measuring the light emission generated by the bound alkaline phosphatase. Light emission is directly proportional to the amount of HIV-1 RNA present in each sample. Each HIV-1 RNA run includes a set of standards of known HIV-1 concentration, allowing the determination of the RNA concentration in the clinical specimens. The dynamic range of this assay is 75 to 500,000 copies/mL. In addition, bDNA is highly reliable in accurately quantifying all subtypes of HIV-1 because the assay includes multiple specific probes targeting the polymerase gene that hybridize to diverse groups of viral genomes. Failure of some of these probes to hybridize to the target sequence has little or no impact on overall signal generation *(18,19)*. A significant portion of the assay may be automated using the Bayer System 340 bDNA Analyzer.

3.4. NUCLISENS HIV-1 RNA QT ASSAY Quantitation with the NucliSens HIV-1 RNA QT ASSAY (bioMerieux Inc., Durham, NC) is based on coamplification of a patient's HIV-1 RNA sample together with internal calibrators. The quantity of amplified RNA is measured by means of electrochemiluminescence (ECL). The NucliSens assay could include an automated nucleic acid extractor that is based on adding lysis

buffer, containing guanidine thiocyanate and Triton X-100, to the clinical sample *(20)*; however, nucleic acid extraction can be performed manually. Three synthetic RNAs of known copy number, high ($Qa10^{4.8}$), medium ($Qb10^{4.2}$), and low ($Qc10^{3.6}$), are added to the lysis buffer containing the viral RNA. These RNAs serve as internal calibrators, each differing from the HIV-1 wild-type (WT) RNA sequence by only a small number of bases, which allows for independent quantification and detection. The extracted nucleic acid, which includes the calibrators, is then subjected to amplification. The HIV-1 RNA and the calibrators serve as targets for the primer that hybridizes to the *gag* region of HIV. The 5′ end of the primer also contains the T7-RNA polymerase recognition site. Avian Myeloblastosis Virus Reverse Transcriptase (AMV-RT) extends the primer and generates a DNA/RNA complex. Following the extension, enzyme RNase H degrades the RNA template and a second primer hybridizes to the cDNA. Synthesis of the second DNA strand is attained through extension of this primer by AMV-RT. Finally, the enzyme T7-RNA polymerase binds to the T7 promoter site and synthesizes multiple copies of the target RNAs. With this RNA synthesis, the system enters the isothermal cyclic phase, resulting in the accumulation of WT and calibrator RNA amplification products. To detect the amplification products (HIV RNA, Qa, Qb, and Qc), aliquots of the amplified sample are added to four hybridization solutions, each with a specific probe for one of the amplification products. The respective amplification products are hybridized with an oligonucleotide fixed to a magnetic bead and a ruthenium-labeled probe. The magnetic beads carrying the hybridized amplification product–probe complex are captured on the surface of an electrode by means of a magnet. Voltage applied to this electrode triggers the electrochemiluminescent reaction. The light emitted by the hybridized ruthenium-labeled probes is proportional to the amount of amplification product. Comparative studies of different subtypes of HIV with this assay have shown that NucliSens is reliable for detecting most subtypes of HIV except quantification of subtypes A and C was significantly less than the expected value *(20)*. The linear range of the assay is 176 to 3.4×10^6 copies/mL. Increasing the volume of plasma used in the NASBA assay improves the sensitivity of the assay *(21)*; the use of 2 mL of plasma will further lower the limit of quantification to approx 80 copies/mL of plasma. The NASBA assay is the only commercial assay designed to accommodate different specimen volumes.

To further automate the NucliSens assay, the real-time detection of amplification products can be achieved through the use of molecular beacon probe technology. Molecular beacons are single-stranded oligonucleotide probes that are in the hairpin form because of self-complementary regions at the 5′ and 3′ end of the probes. The probe contains a fluorescent molecule on one arm and a quencher molecule on the other arm. When the probe is in the hairpin form, it does not fluoresce because of proximity of the fluorophore and the quencher. When the probe binds to its specific target, the quencher is spatially separated from the fluorophore, which then generates fluorescence. To utilize this system for the detection of RNA products, the NASBA amplification mix contains a single calibrator and two molecular beacon probes: one for HIV-target amplification products and one for calibrator amplification products. The two probes have distinct fluorescent dyes and generate specific fluorescence signals when annealed to the amplification products. The fluorescent signal is monitored in real time without opening the amplification vial. In addition to real-time detection of amplification products, the molecular-beacon-based detection method virtually eliminates the amplification product carryover contamination that is known to lead to false-positive results *(22,23)*.

3.5. ROCHE AMPLICOR HIV MONITOR™ V1.5 ASSAY

The Amplicor HIV Monitor™ Assay (Roche Diagnostics, Indianapolis, IN) is a RT-PCR-based viral-load assay. The plasma RNA is extracted with a lysis reagent containing guanidine thiocyanate and quantitation standard (QS) RNA. The QS that is added into each sample permits quantification of HIV-1 RNA by comparison of resulting optical densities following amplification and detection. A 142-bp sequence in the *gag* gene of HIV and the QS (primers SK145 and SKCC1B) are amplified by RT-PCR in a single reaction. The primers are biotinylated at the 5 ends and the HIV-1 and QS amplicons are detected in separate wells of a microwell plate (MWP) coated with HIV-specific and QS-specific oligonucleotide capture probes, respectively. The bound, biotinylated amplicons are quantified with an avidin–horseradish peroxidase (HRP) conjugate and a colorimetric reaction for HRP. The HIV-1 RNA copy number is then calculated from the known input copy number of the QS RNA. The reported dynamic range of the standard assay is 400 to 750,000 copies/mL. The lower limit of the assay is further reduced in an ultrasensitive version of the test. The ultrasensitive HIV Monitor assay requires initial ultracentrifugation of the sample for 1 h prior to RNA extraction and has a dynamic range of 50 to 75,000 viral particles/mL (the version 1.5 assay has an upper limit of sensitivity of 100,000 copies/mL) *(24)*. In comparative studies of different subtypes of HIV-1, the version 1.0 Amplicor assay was not reliable for detecting HIV-1 subtypes A, E, and F *(25)*. Because of this limitation of the version 1.0 assay, version 1.5 of the Amplicor assay has been developed with modified primers that allow for the accurate quantification of subtypes A–H *(24,25)*. The version 1.5 assay has recently been FDA approved and has replaced the version 1.0 assay in clinical practice.

The Amplicor assays are available in either a MWP format or on the Cobas platform, which automates the amplification and detection steps of the assay. In addition, the use of the MagNA Pure LC DNA extraction system (Roche Diagnostics) provides further automation of the nucleic acid extraction step.

3.6. TAQMAN The next generation of PCR-based techniques for monitoring HIV-1 will utilize real-time RT- PCR (TaqMan) technologies *(26,27)*. Similar to the molecular beacons, TaqMan assays allow for amplification and detection to be performed in a single closed tube, virtually eliminating amplification product carryover contamination. The principle of TaqMan real-time PCR is based on the cleavage of an internal probe by the 5′–3′ exonuclease activity of *Taq* polymerase during amplification. The probe contains a fluorescent reporter and a quencher molecule at its 5′ and 3′ ends, respectively. At the start of the reaction there is no fluorescence because of the proximity of quencher and reporter dye. If the probe is hybridized

to the target, the hybridized probe will be cleaved by the 5′ to 3′ nuclease activity of the *Taq* DNA polymerase and the reporter dye will be separated from the quencher dye, leading to generation of fluorescent signal during each cycle of PCR. One molecule of reporter dye is released for each target molecule amplified. A passive reference dye is included to provide an internal reference to normalize for well-to-well variation of factors such as illumination. The threshold cycle (Ct) value is the number of PCR cycles before fluorescence emitted passes a fixed limit. The $\log_{10}$ of the number of targets initially present is proportional to Ct and can be measured using a standard curve *(26)*.

Results between the different viral-load assays might not be comparable because each uses a different molecular technique and might use a different HIV-1 RNA standard. However, the viral-load levels obtained with the bDNA (version 3.0) and the Amplicor RT-PCR assay (version 1.5) were highly correlated in one study and differences between the two assays were within the coefficients of variation for the assays *(19)*. The Amplicor version 1.0 and 1.5 assays have been reported to generate comparable quantification on matched clinical specimens *(24,28)*. Although the differences in viral-load levels obtained with the different assays are narrowing, it is still best to use the same assay when monitoring patients over time.

False-positive results can occur with viral-load assays and are attributed to limitations in assay chemistry, carryover contamination with amplicons, or contamination during specimen processing. The bDNA assay chemistry involves complex hybridization of nucleic acid probes with signal amplification. This assay has a specificity of approx 98% when testing specimens from HIV-1-negative individuals *(29)*. Carryover contamination of amplified products does not occur with the bDNA assay, but can lead to false-positive results with the NASBA and Amplicor RT-PCR assays. For the Amplicor RT-PCR assay, the risk of amplicon contamination is reduced because UTP and the DNA repair enzyme uracil-*N*-glycosylase (UNG) are used. Contamination with HIV-1 RNA during specimen processing can lead to false-positive results with any of the three assays.

Each of the viral-load assays has their strengths and weaknesses. The Versant bDNA assay has the advantages of high throughput, a broad dynamic range, and the ability to accurately quantify all subtypes of HIV-1. However, the assay does not have an internal control to evaluate the quality of nucleic acid extraction and false-positive results have been reported *(29)*. The Amplicor RT-PCR assay has good specificity, but the dynamic range is limited, requiring both a standard and ultrasensitive version of the assay. In addition, the version 1.0 Amplicor RT-PCR assay underestimated some non-B subtypes, although this problem has been resolved now that version 1.5 has become available. The NASBA assay has a broad dynamic range and can accommodate a wide range of clinical specimen types and specimen volumes, but it underestimates some non-B subtypes. All three assays have automated or semiautomated platforms available or in development. The choice of assays for a specific laboratory will depend on several factors, including the expertise of the technologists, available space, volume of testing, turnaround time required, and cost of testing.

Proficiency testing for HIV-1 viral-load measurements is available from the College of American Pathologists Surveys program (CAP; HIV/HV2 survey). The CAP program offers testing three times a year. The Centers for Disease Control and Prevention (CDC) offers laboratory performance evaluations twice a year through the CDC Model Performance Evaluation Program. In addition, several companies, including AcroMetrix (Benicia, CA), Boston Biomedica, Inc. (West Bridgewater, MA), and IMPATH-BCP, Inc. (Franklin, MA), provide external standards and verification panels for HIV-1 viral-load testing. In addition, the World Health Organization (WHO) has also established an HIV-1 RNA reference standard *(30)*.

3.7. SPECIMEN REQUIREMENTS Plasma specimens are most commonly used for viral-load assays. The volume of plasma used varies with the different assays: the standard version of the Amplicor RT-PCR Monitor assay uses 0.2 mL of plasma, the ultrasensitive Amplicor RT-PCR Monitor assay utilizes 0.5 mL of plasma, and the Versant bDNA assay (version 3.0) utilizes 1.0 mL of plasma. The standard volume for the NucliSens NASBA assay is 1.0 mL of plasma, although volumes between 50 µL and 2 mL can be used. To ensure accurate HIV-1 RNA quantification, proper specimen collection and processing are essential in order to minimize RNA degradation. For the Versant bDNA and Amplicor RT-PCR assays, EDTA is the preferred anticoagulant, whereas blood anticoagulated with heparin is unacceptable. Blood anticoagulated with acid citrate dextrose (ACD, yellow top tube) is acceptable, but viral load will be decreased approx 15% as a result of the volume of anticoagulant *(31–33)*. The NucliSens NASBA assay can accommodate any anticoagulant because of the extraction method that is used *(34)*. For all assays, it is recommended that plasma be separated within 4–6 h of collection, as delays in separation could lead to a falsely decreased viral load resulting from RNA degradation, although more recent studies have shown that plasma specimens can be stored at 4°C for several days without significant degradation of RNA. Moreover, HIV-1 RNA remains stable after three cycles of freezing (–70°C) and thawing *(33)*. For long-term storage, plasma samples should be frozen at or below –70°C *(32)*. Vacutainer plasma preparation tubes (PPTs) are a convenient alternative for the collection of blood specimens for viral-load testing. PPTs contain a gel barrier, which, after centrifugation, physically separates plasma from the cellular components *(31)*. This process stabilizes HIV-1 RNA levels for up to 30 h after collection in PPTs when stored at either room temperature or at 4°C as cell free plasma or as EDTA-anticoagulated blood *(31)*. PPTs allow for a closed sample collection system that is safe and convenient.

3.8. INTERPRETATION OF VIRAL-LOAD MEASUREMENTS The HIV-1 viral load assays have low intra-assay variability (approx 0.12–0.2 $\log_{10}$) on repeated testing of single samples *(32,35)*, with the bDNA assay showing the lowest values. For all assays, variability is greater near the lower limit of sensitivity, so it is important not to overinterpret small changes in viral load, particularly with the ultrasensitive assays. HIV-l RNA levels are relatively stable in untreated individuals *(17)* with a biological variation of approx 0.3 $\log_{10}$. Therefore, changes in viral load must be at least three-fold (0.5 $1\log_{10}$) in magnitude in order to represent biologically relevant changes in

the level of viral replication. There are a variety of clinical syndromes that can cause a transient increase in viral load, including herpes simplex virus (HSV) infections, opportunistic infections, and acute respiratory infections. In addition, vaccinations for influenza, tetanus, or pneumococcal infection transiently increase viral-load levels *(36–38)*. For some patients, the increase in viral load could be quite dramatic, greater than or equal to 1 $\log_{10}$. As results usually return to baseline within 1 mo, HIV-l RNA levels should not be measured within a month of any of these events.

Although not FDA approved as a diagnostic test, viral-load testing can be useful in the diagnosis of acute HIV-1 infection *(29)*. Acute HIV infection refers to the "window period" after exposure and prior to seroconversion, when the enzyme-linked immunosorbent assay (ELISA) and Western blot assay are negative or indeterminate. However, the amount of virus in the plasma of patients with acute HIV-1 infection is very high (usually 10^5–10^7 copies/mL plasma) *(29)*. Viral-load testing has been shown to be more sensitive than p24 testing for identifying patients with acute HIV-1 infection; however, p24 antigen testing is more specific *(29)*. The detection of proviral DNA is another option for the diagnosis of acute HIV-1 infection. However, several studies have shown that HIV-1 RNA testing is more sensitive than proviral DNA testing for the diagnosis of acute HIV-1 infection in newborns *(39,40)*. The use of molecular assays or p24 antigen testing in patients with suspected acute HIV-1 infection should be performed with caution. Patients must be educated about the limitations of these tests and give informed consent prior to testing. An HIV-1/2 ELISA should also be obtained at the time of viral-load testing, and repeat testing should be done on all specimens with a detectable viral load.

4. HIV-1 RESISTANCE TESTING

The use of combination therapy is now the standard of care when treating HIV-1-infected individuals, as this approach delays the development of resistant virus. Patients are now routinely treated with combinations of nucleoside reverse transcriptase inhibitors and either a non-nucleoside reverse transcriptase inhibitor or a protease inhibitor. In addition to these three classes of drugs, the first member of a new class of drug, the fusion inhibitors, is now available. However, in spite of the use of combination therapy and the increasing number of antiretroviral drugs that are now available, the development of resistant virus is an important cause of treatment failure. As a result, resistance testing is now used routinely in the management of HIV-1-infected individuals.

In 2000, the International AIDS Society–USA panel recommended the use of resistance testing to help guide the choice of new anti-HIV regimens following treatment failure and for guiding therapy for pregnant women. More recent guidelines *(41)* recommend resistance testing for persons failing therapy, for pregnant women, and for persons who have been infected within the past 2 yr. With the use of multiple-drug therapy to treat HIV infection, resistance testing can be used to identify which of the administered drugs in a regimen are likely to be ineffective. The panel also recommended that testing be considered prior to initiating therapy in patients with acute HIV infection, because in some individuals the primary HIV infection is with resistant virus. Several reports have concluded that resistance testing could be cost-effective. Chaix et al. reported that the cost of drug resistance testing in their patients was offset by a reduced use of protease inhibitors *(42)*. In a recent study, Weinstein et al. concluded that genotyping following antiviral failure is cost-effective and that resistance testing for treatment of naive individuals might be reasonable as an increasing number of individuals become infected with resistant virus *(43)*.

Both genotypic and phenotypic resistance assays are used in clinical practice. Genotypic resistance assays detect specific mutations in the viral genome that are associated with resistance to various antiretroviral agents, whereas phenotypic assays measure viral replication in the presence of antiviral drugs. For genotypic assays, the initial step is RT-PCR amplification of viral protease gene and a 250- to 350-codon segment of the HIV RT gene. This is followed by either direct sequencing of the amplified products or by using reverse-hybridization technology. The majority of genotypic testing done for clinical management relies on automated sequencing technology. This approach is quite complex, requiring sequence alignment, sequence editing, and comparison of data to WT sequence to allow for identification of the specific resistance mutations. Finally, an interpretation of the data is required, which maps the mutations to specific antiretroviral drugs. Currently, there are two FDA-cleared sequencing assays (Trugene HIV-1 Genotyping Kit and OpenGene DNA Sequencing System, Bayer Diagnostics, Tarrytown, NY; and ViroSeq HIV-1 Genotyping System, Abbott Diagnostics, Abbott Park, IL), which provide reagents as well as the software programs to assist with sequence identification and interpretation. In addition, several investigators have described in-laboratory-developed sequencing assays.

Phenotypic assays measure viral replication in the presence of various antiretroviral drugs. Currently available phenotypic assays rely on recombinant technology, which allows more rapid testing than traditional methods that required isolation of infectious virus. Viral replication is measured in varying concentrations of drug, allowing for the determination of the amount of drug required to inhibit viral replication by 50% (IC_{50}). The IC_{50} of the patient strain is compared to that of a WT virus and fold differences in IC_{50} are reported. Phenotypic testing is performed in two commercial laboratories (ViroLogic, South San Francisco, CA and Viroc, Mechelen, Belgium). A "virtual phenotype" is an alternative to phenotypic testing in which the genotype results are entered into a database containing thousands of matching genotype and phenotype results. The virtual phenotype is then determined by averaging the closest matching phenotype results.

Genotypic assays are available in clinical laboratories and are technically easier to perform, have a more rapid turnaround time, and are less expensive than phenotypic assays. However, the interpretation of genotypic assays is complex and requires an extensive knowledge of the genetics of resistance, including issues of cross-resistance and interaction of mutations. Several drug manufacturers and independent investigators have developed Internet-based sites with large databases that have, to a degree, alleviated this problem *(44,45)*. These sites are free of charge and are frequently updated to include the most recent

published data. Phenotypic assays provide a direct measure of susceptibility but require 1–2 wk to perform, and clinically relevant changes in IC_{50} need to be established for each drug. Moreover, the detection of drug mutations might precede a change in susceptibility, so genotypic and phenotypic results might not always correlate. There is no standard of care regarding whether to use genotypic or phenotypic resistance testing, and depending on the clinical situation, both tests have utility.

5. SUMMARY

The development of the HIV pandemic has in an unprecedented manner allowed for the simultaneous introduction and development of therapeutics and molecular diagnostics. Molecular methods have rapidly become indispensable scientific and diagnostic tools that are now considered routine in most clinical laboratories. The use of viral-load and resistance testing clearly underscores the critical role that molecular diagnostics plays in clinical care.

REFERENCES

1. Prusiner, S. B., Historical essay. Discovering the cause of AIDS. [comment]. *Science* **298(5599):**1726, 2002.
2. Berry, N. and Tedder, R. HIV-1 and HIV-2 molecular diagnosis, in *HIV and New Viruses*, R. A. Dalgleish, ed., Academic, San Diego, CA, pp. 207–222, 1999.
3. Fitzgibbon, J. E., Howell, R. M., Haberzettl, C. A., Sperber, S. J., Gocke, D. J., and Dubin, D. T. Human immunodeficiency virus type 1 pol gene mutations which cause decreased susceptibility to 2¢,3¢-dideoxycytidine. *Antimicrob. Agents Chemother.* **36(1):**153–157, 1992.
4. Larder, B. A., Darby, G., and Richman, D. D. HIV with reduced sensitivity to zidovudine (AZT) isolated during prolonged therapy. *Science* **243(4899):**1731–1734, 1989.
5. Shafer, R. W., Kozal, M. J., Winters, M. A., et al. Combination therapy with zidovudine and didanosine selects for drug-resistant human immunodeficiency virus type 1 strains with unique patterns of pol gene mutations. *J. Infect. Dis.* **169(4):**722–729, 1994.
6. Tisdale, M., Myers, R. E., Maschera, B, Parry, N. R., Oliver, N. M., and Blair, E. D. Cross-resistance analysis of human immunodeficiency virus type 1 variants individually selected for resistance to five different protease inhibitors. *Antimicrob. Agents Chemother.* **39(8):**1704–1710, 1995.
7. Shirasaka, T., Yarchoan, R., O'Brien, M. C., et al. Changes in drug sensitivity of human immunodeficiency virus type 1 during therapy with azidothymidine, dideoxycytidine, and dideoxyinosine: an in vitro comparative study. *Proc. Natl. Acad. Sci. USA* **90(2):**562–566, 1993.
8. Little, S. J., Daar, E. S., D'Aquila, R. T., et al. Reduced antiretroviral drug susceptibility among patients with primary HIV infection. [see comment]. *J. Am. Med. Assoc.* **282(12):**1142–1149, 1999.
9. Little, S. J. Transmission and prevalence of HIV resistance among treatment-naive subjects. *Antiviral Ther.* **5(1):**33–40, 2000.
10. Weiss, R. A. Gulliver's travels in HIVland. *Nature,* **410(6831):** 963–967, 2001.
11. Mellors, J. W., Kingsley, L. A., Rinaldo, C. R., Jr., et al. Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion. *Ann. Intern. Med.* **122(8):**573–579, 1995.
12. Mellors, J. W., Rinaldo C. R., Jr., Gupta, P., White, R. M., Todd, J. A., and Kingsley, L. A. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science,* **272:**1167–1170, 1996.
13. Cameron, D. W., Heath-Chiozzi, M., Danner, S., et al. Randomised placebo-controlled trial of ritonavir in advanced HIV-1 disease. The Advanced HIV Disease Ritonavir Study Group. *Lancet* **351(9102):** 543–549, 1998.
14. Ho, D. D., Neumann, A. U., Perelson, A. S., Chen, W., Leonard, J. M., and Markowitz, M. Rapid turnover of plasma virions and CD4+ lymphocytes in HIV-1 infection. *Nature* **373:**123–126, 1995.
15. Wei, X., Ghosh, S. K., Taylor, M. E., et al. Viral dynamics in human immunodeficiency virus type 1 infection. *Nature* **373:**117–122, 1995.
16. Raboud, J. M., Montaner, J. S., Conway, B., et al. Suppression of plasma viral load below 20 copies/ml is required to achieve a long-term response to therapy. *AIDS* **12(13):**1619–1624, 1998.
17. Saag, M. S., Holodniy, M., Kuritzkes, D. R., et al. HIV viral load markers in clinical practice. *Nat. Med.* **2:**625–629, 1996.
18. Erice, A., Brambilla, D., Bremer, J., et al. Performance characteristics of the QUANTIPLEX HIV-1 RNA 3.0 assay for detection and quantitation of human immunodeficiency virus type 1 RNA in plasma. *J. Clin. Microbiol.* **38(8):**2837–2845, 2000.
19. Elbeik, T., Charlebois, E., Nassos, P., et al. Quantitative and cost comparison of ultrasensitive human immunodeficiency virus type 1 RNA viral load assays: Bayer bDNA quantiplex versions 3.0 and 2.0 and Roche PCR Amplicor Monitor version 1.5. *J. Clin. Microbiol.* **38(3):** 1113–1120, 2000.
20. Ginocchio, C., Kemper, M., Stellrecht, K., and Witt, D. J. Multicenter evaluation of the performance characteristics of the NucliSens HIV-1 QT assay used for quantification of human immunodeficiency virus type 1 RNA. *J. Clin. Microbiol.* **41(1):** 164–173, 2003.
21. Witt, D. J., Kemper, M., Stead, A., Ginocchio, C., and Caliendo, A. M. Relationship of incremental specimen volumes and enhanced detection of human immunodeficiency virus type 1 RNA with nucleic acid amplification technology. *J. Clin. Microbiol.* **38:**85–89, 2000.
22. Niesters, H. G. Quantitation of viral load using real-time amplification techniques. *Methods* **25(4):**419–429, 2001.
23. Tyagi, S. and Kramer, F. R. Molecular beacons: probes that fluoresce upon hybridization. *Nat. Biotechnol.* **14(3):**303–308, 1996.
24. Triques, K., Coste, J., Perret, J. L., et al. Efficiencies of four versions of the AMPLICOR HIV-1 MONITOR test for quantification of different subtypes of human immunodeficiency virus type 1. *J. Clin. Microbiol.* **37(1):**110–116, 1999.
25. Jagodzinski, L. L., Wiggins, D. L., McManis, J. L., et al. Use of calibrated viral load standards for group M subtypes of human immunodeficiency virus type 1 to assess the performance of viral RNA quantitation tests. *J. Clin. Microbiol.* **38(3):**1247–1249, 2000.
26. de Baar, M. P., Timmermans, E. C., Bakker, M., de Rooij, E., van Gemen, B., and Goudsmit, J. One-tube real-time isothermal amplification assay to identify and distinguish human immunodeficiency virus type 1 subtypes A, B, and C and circulating recombinant forms AE and AG. *J. Clin. Microbiol.* **39(5):**1895–1902, 2001.
27. Wittwer, C. T., Herrmann, M. G., Moss, A. A., and Rasmussen, R. P. Continuous fluorescence monitoring of rapid cycle DNA amplification. *Biotechniques* **22(1):**130–131, 134–138, 1997.
28. Hill, C. E., Green, A. M., Ingersoll, J., Easley, K. A., Nolte, F. S., and Caliendo, A. M. Assessment of agreement between the AMPLICOR HIV-1 MONITOR test versions 1.0 and 1.5. *J. Clin. Microbiol.* **42(1):**286–289, 2004.
29. Daar, E. S., Little, S., Pitt, J., et al. and Los Angeles County Primary H. I. V. Infection Recruitment Network. Diagnosis of primary HIV-1 infection. Los Angeles County Primary HIV Infection Recruitment Network. *Ann. Intern. Med.* **134(1):**25–29, 2001.
30. Holmes, H., Davis, C., Heath, A., Hewlett, I., and Lelie, N. An international collaborative study to establish the 1st international standard for HIV-1 RNA for use in nucleic acid-based techniques. *J. Virol. Methods* **92(2):**141–150, 2001.
31. Ginocchio, C. C., Wang, X. P., Kaplan, M. H., et al. Effects of specimen collection, processing, and storage conditions on stability of human immunodeficiency virus type 1 RNA levels in plasma. *J. Clin. Microbiol.* **35(11):**2886–2893, 1997.
32. Lew, J., Reichelderfer, P., Fowler, M., et al. Determinations of levels of human immunodeficiency virus type 1 RNA in plasma: reassessment of parameters affecting assay outcome. TUBE Meeting Workshop Attendees. Technology Utilization for HIV-1 Blood Evaluation and Standardization in Pediatrics. *J. Clin. Microbiol.* **36:**1471–1479, 1998.
33. Sebire, K., McGavin, K., Land, S., Middleton, T., and Birch, C. Stability of human immunodeficiency virus RNA in blood specimens as measured by a commercial PCR-based assay. *J. Clin. Microbiol.* **36(2):**493–498, 1998.

34. Boom, R., Sol, C., Salimans, M., Jansen, C., Wertheim-van Dillen, P. M., and van derNoordaa, J. Rapid and simple method for purification of nucleic acids. *J. Clin. Microbiol.* **28:**495–503, 1990.
35. Brambilla, D., Leung, S., Lew, J., et al. Absolute copy number and relative change in determinations of human immunodeficiency virus type 1 RNA in plasma: effect of an external standard on kit comparisons. *J. Clin. Microbiol.* **36(1):**311–314, 1998.
36. Donovan, R. M., Bush, C. E., Markowitz, N. P., Baxa, D. M., and Saravolatz, L. D. Changes in virus load markers during AIDS-associated opportunistic diseases in human immunodeficiency virus-infected persons. *J. Infect. Dis.* **174(2):**401–403, 1996.
37. O'Brien, W. A., Grovit-Ferbas, K., Namazi, A., et al. Human immunodeficiency virus-type 1 replication can be increased in peripheral blood of seropositive patients after influenza vaccination. *Blood* **86(3):**1082–1089, 1995.
38. Staprans, S. I., Hamilton, B. L., Follansbee, S. E., et al. Activation of virus replication after vaccination of HIV-1-infected individuals. *J. Exp. Med.* **182(6):**1727–1737, 1995.
39. Cunningham, C. K., Charbonneau, T. T., Song, K., et al. Comparison of human immunodeficiency virus 1 DNA polymerase chain reaction and qualitative and quantitative RNA polymerase chain reaction in human immunodeficiency virus 1-exposed infants. *Pediatr. Infect. Dis. J.* **18(1):**30–35, 1999.
40. Steketee, R. W., Abrams, E. J., Thea, D. M., et al. and the New York City Perinatal HIV Transmission Collaborative Study. Early detection of perinatal human immunodeficiency virus (HIV) type 1 infection using HIV RNA amplification and detection. *J. Infect. Dis.* **175:**707–711, 1997.
41. Hirsch, M. S., Conway, B., D'Aquila, R. T., et al. Antiretroviral drug resistance testing in adults with human immunodeficiency virus type 1: 2003 recommendations of an international AIDS Society-USA panel. *Clin. Infect. Dis.* **37:**113–128, 2003.
42. Chaix, C., Grenier-Sennelier, C., Clevenbergh, P., Schapiro, J. M., Dellamonica, P., and Durand-Zaleski, I. Economic evaluation of drug resistance genotyping for the adaptation of treatment in HIV-infected patients in the VIRADAPT study. *J. AIDS* **24(3):** 227–231, 2000.
43. Weinstein, M. C., Goldie, S. J., Losina, E., et al. Use of genotypic resistance testing to guide HIV therapy: clinical impact and cost-effectiveness. *Ann. Intern. Medi.* **134(6):**440–450, 2001.
44. Shafer, R. W., Stevenson, D., and B. Chan. Human Immunodeficiency Virus Reverse Transcriptase and Protease Sequence Database. *Nucleic Acids Res,* **27(1):**348–352, 1999.
45. Shafer, R. W., Jung, D. R., Betts, B. J., Xi, Y., and Gonzales, M. J. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.* **28(1):**346–348, 2000.

37 Hepatitis C

D. ROBERT DUFOUR

1. INTRODUCTION

Hepatitis C virus (HCV) is the most common chronic viral infection in North America and Europe and a common infection worldwide. In the United States Third National Health and Nutrition Examination Survey, it was estimated that 3.9 million people had detectable antibodies to HCV, and 75% of them were positive for HCV RNA *(1)*. This survey included only individuals living in households. There is evidence that HCV is much more common in certain populations, such as the homeless and prison inmates, where prevalence might be as high as 40% *(2)*. The prevalence of HCV is also higher among veterans seeking care in VA medical centers; nationwide, the prevalence is just under 7% *(3)*, with a prevalence over 10% in urban hospitals *(4)*. Worldwide, an estimated 170 million individuals are chronically infected with HCV *(5)*. The age of infected individuals differs significantly in different areas of the world. In Australia, North America, and much of Europe, infections seems to peak in persons currently between the ages of 40 and 60, whereas in Japan, Italy, and Egypt, prevalence of HCV infection increases steadily with increasing age *(6)*. Males are more likely to be infected than females *(1)*. The prevalence of acute infections with HCV has fallen by 80% over the past decade *(7)*.

There are a number of recognized risk factors for HCV. The most common are injection drug use and transfusion or transplantation before 1992 *(8)*. Although the overwhelming majority of those using injection drugs in the 1960s and 1970s became infected with HCV *(9)*, the likelihood of infection among current, younger injecting drug users is only about 50% *(10)*, this reflects an overall decline in the incidence of acute HCV infection among injecting drug users *(7)*. Even among blood donors who initially deny risk factors but are found to be HCV positive, a history of injection drug use on follow-up questioning is 50 times more likely than in donors who are HCV negative *(11)*. Less commonly, HCV is transmitted by dialysis *(12,13)*, by needle-stick injury *(14)*, and by vertical transmission from an infected mother to her child *(15–17)*; the likelihood of infection with the last two of these routes is approx 3–5%. The role of sexual transmission is controversial, but evidence suggests that having multiple sexual partners increases risk of transmission of HCV *(18–21)*; the likelihood of infection in monogamous sex partners of HCV-infected individuals is similar to that in the population as a whole, however *(22,23)*.

Laboratory tests are the major means to diagnose and monitor HCV infection. This chapter will review the principles behind laboratory tests for HCV, as well as their uses in clinical practice. It will also briefly discuss potential uses for molecular tests in future evaluation of those affected by HCV.

2. VIRAL STRUCTURE AND FUNCTION

Hepatitis C virus has never been isolated in culture. It was recognized in 1989 by using recombinant technology to create peptides from an infectious serum that were tested against serum from individuals with non-A, non-B hepatitis, and a section of the viral genome was isolated *(24)*. Ultimately, the entire HCV genome was sequenced *(25)*. Hepatitis C virus is a member of the family of flaviviridae; other members of the family include yellow fever virus, West Nile virus, and hepatitis GB virus. Flaviviruses are positive, single-stranded RNA viruses. Hepatitis C virus codes for production of a single polypeptide chain of approx 3000 amino acids. The peptide codes for a number of specific proteins, as illustrated in Fig. 1.

A number of the proteins of HCV appear to be important in HCV function or in HCV treatment. The Env proteins are among the most variable parts of the peptide chain and are associated with multiple molecular forms in a single infected person *(26,27)*. These mutations seem to be critical for escape from the host immune response *(28)*. The NS5a protein contains an interferon-response element; there is evidence in many studies that mutations in this gene are associated with resistance to interferon, the main antiviral agent used in the treatment of HCV *(29)*. The NS3 region codes for a protease, whereas the NS5b codes for an RNA polymerase. Drugs that target protease or polymerase are now undergoing trials as therapeutic agents to treat HCV infection *(30)*.

Hepatitis C viral strains differ significantly from each other. The nomenclature adopted has divided HCV RNA into three major levels: genotypes, subtypes, and quasispecies. There are six genotypes of HCV, numbered from 1 to 6; each genotpye has <70% homology in the nucleotide sequence compared to the others. Subtypes typically have 77–80% homology in nucleotide sequence, whereas quasispecies have over 90% homology *(31)*.

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

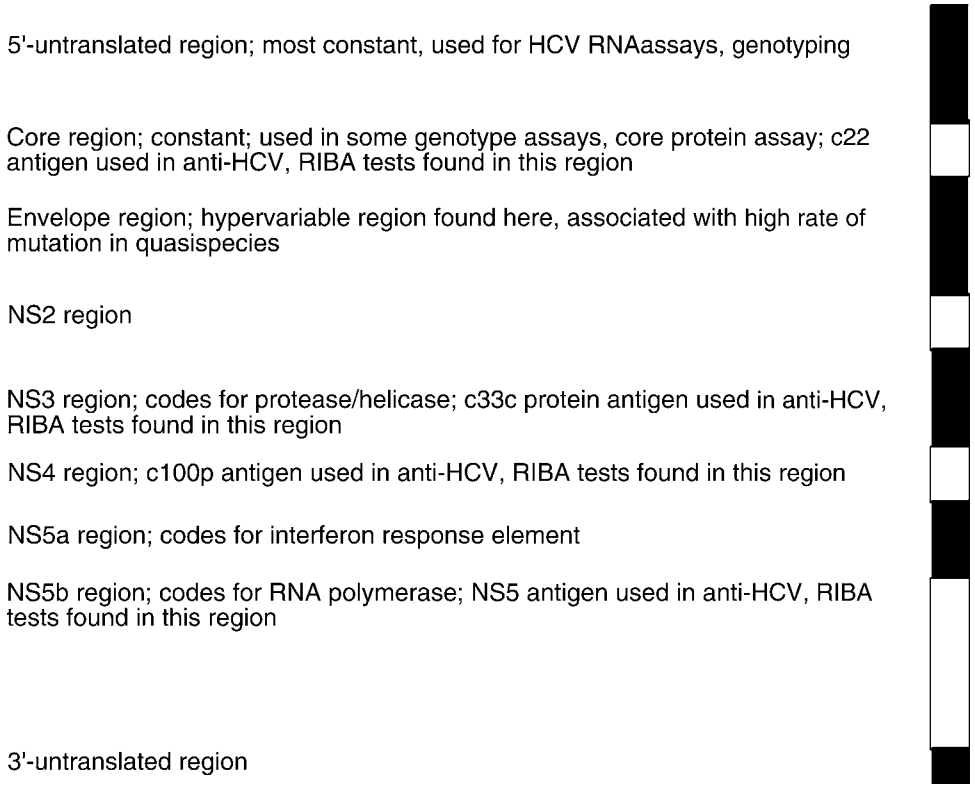


Fig. 1. Structure–function correlates in hepatitis C. This figure illustrates the relative size and function of different regions of the HCV genome, from the 5′ (top) to the 3′ (bottom) end of the viral RNA.

Infections are of a single genotype and subtype (except in rare instances), but each individual has many quasispecies of HCV.

Because RNA viruses do not contain a proof-reading mechanism, mutations in the genome are common. In the case of HCV, mutations seem to emerge after the development of antibodies during the period of initial infection. If mutant strains do not emerge, there is a high likelihood that the virus will be cleared from the infected individual *(32)*. Eventually, a given individual will have a large number of quasispecies of HCV in their liver and bloodstream. The pattern of mutation is not reproducible from individual to individual; several studies have shown divergence over time in the pattern of quasispecies in the donor and recipient of a single strain of HCV infection *(33,34)*.

Although there is a high degree of variability in nucleotide sequence between different genotypes, there are certain parts of the HCV genome that show little difference between different genotypes or quasispecies; these include the 5′-untranslated region (the most highly conserved region), the core region, and the NS5b region *(31)*. Tests for HCV RNA and genotyping typically utilize the 5′-untranslated region and/or the core region to minimize the effects of quasispecies and genotype differences on test performance. The core antigen has been used as a diagnostic test for HCV infection.

2.1. CLINICAL ASPECTS OF HEPATITIS C INFECTION

Hepatitis C seems to infect predominantly liver cells, although there is evidence that infection also occurs in other cells, particularly circulating mononuclear cells *(35)*. The primary manifestations of HCV infection are, therefore, those involving the liver. With initial infection with HCV, there is a latent period of 2–4 wk before viral replication is detectable *(36)*. In most cases, even after viremia develops, there is no clinical evidence of the infection; it is estimated that only 10–30% of those with infection develop a picture of acute hepatitis *(37)*. When acute hepatitis develops, it can be clinically manifested by fever, loss of appetite, nausea, diarrhea, and specific liver symptoms such as discomfort and tenderness in the right upper abdomen, jaundice, dark urine, and pale colored stools. Typically, this occurs from 2 to 3 mo after initial infection and gradually resolves over a period of several weeks *(38)*. During this time, liver enzymes such as alanine and aspartate aminotransferases (ALT and AST, respectively) are typically increased to 10–40 times above the upper reference limit *(39)*. Data suggest that those who develop a symptomatic acute infection are more likely to clear the virus than those who do not *(40,41)*, although not all studies have noted such a correlation *(42)*.

In most individuals infected with HCV, there are no signs or symptoms that accompany the initial infection. In most cases, a chronic infection with HCV develops, usually producing chronic hepatitis, a state of ongoing inflammation in the liver. Chronic HCV infection is typically clinically silent for many years or produces only mild, nonspecific symptoms such as fatigue, loss of energy, and difficulty concentrating. Most persons with chronic hepatitis resulting from HCV have mildly increased activity levels of ALT and, less commonly, AST; activities typically fluctuate over time, and in the majority of individuals, they vary between increased and normal *(43)*. Peak ALT activity levels are usually between one and four times the upper reference limit, and about 10–15% of individuals with

Table 1
Factors Associated with Increased Risk of Fibrosis in Chronic Hepatitis C

Age > 40 yr at time of infection
Male gender
Caucasian race
Transfusion as source of infection
Heavy alcohol intake (>60 g per day)
Immunosuppression
Recurrent infection posttransplantation

chronic infection have persistently normal ALT activity values *(44,45)*. Although there is a correlation between ALT activity and the degree of injury seen in liver biopsies *(46,47)*, individuals with consistently normal ALT might have significant injury and be at risk for progression to more severe complications.

The major end-stage forms of chronic hepatitis are cirrhosis and hepatocellular carcinoma. Cirrhosis, the 10th leading cause of death in the United States *(48)*, is a product of chronic hepatitis and scarring in the liver. It has been estimated that 20–30% of those with chronic HCV will progress to cirrhosis after 20 yr of infection, although fibrosis progresses at different rates in different individuals *(49,50)*. Factors associated with a higher risk of fibrosis progression are summarized in Table 1. Several studies in those infected as children or young adults suggest a much lower rate of progression to cirrhosis after 20 yr *(51–53)*. Once cirrhosis develops, there is often a period in which symptoms are minimal (termed compensated cirrhosis). Clinical signs of decompensation such as ascites, upper gastrointestinal tract bleeding, or hepatic encephalopathy are associated with a 60% risk of death within 5 yr *(54)*, and often lead to consideration for liver transplantation. Cirrhosis resulting from HCV has become the most common indication for liver transplantation in the United States *(2)*.

Hepatocellular carcinoma, a primary tumor of the liver cells, is a relatively uncommon form of cancer in North America and Europe, although its incidence approximately doubled over the past 30 yr *(55)* and is projected to increase another twofold to threefold by 2030 *(56)*. Worldwide, hepatocellular carcinoma is the fourth leading cause of cancer death *(57)*. The major risk factors for development of this cancer are chronic infection with hepatitis B and C viruses; although hepatitis B is the most common risk factor in many parts of the world *(58)*, HCV is responsible for the majority of cases in North America, Europe, and Japan *(59,60)*. Once cirrhosis has developed in HCV, the risk of development of hepatocellular carcinoma has been estimated as 1–4% per year *(59)*.

Although chronic liver disease is the major manifestation of chronic HCV infection, HCV is a common cause of a number of disorders involving other organs *(61)*. In the skin, HCV is a major risk factor for development of porphyria cutanea tarda, in which high levels of porphyrin precursors (notably uroporphyrins) lead to development of blisters in sun-exposed skin. Lichen planus, a chronic inflammatory skin condition, is also more common in HCV. Cryoglobulinemia, characterized by immunoglobulins that precipitate at cold temperatures, is commonly associated with HCV *(62)*. Precipitates of cryoglobulins can obstruct small vessels in the hands, feet, and legs, and in severe cases, they can lead to amputations. Rarely, HCV-related immune complexes cause glomerulonephritis. Malignant lymphomas occur more frequently in those with HCV, and in at least one unusual form of lymphoma, clearance of HCV by antiviral treatment led to remission of the lymphoma *(63)*. HCV has also been linked epidemiologically to type 2 diabetes mellitus *(64)*.

2.2. SEROLOGIC TESTS FOR HEPATITIS C Most testing for HCV begins with detection of antibodies to HCV proteins. The initial test for HCV involved a single HCV antigen (5-1-1); subsequent generations have included antigens from four distinct parts of the HCV polypeptide. Most clinical laboratories still use second-generation HCV antibody assays, whereas most blood banks use third-generation antibody assays. Although the most common format for antibody assays has been enzyme immunoassay, commercially available assays using microparticle immunoassay and chemilluminescence have also been developed.

Most laboratories have reported anti-HCV assay results as positive or negative. Results are interpreted based on a cutoff value provided by the manufacturer; those equal to or above the cutoff are termed positive and those below the cutoff are negative. There is a biphasic distribution among positive results in the ratio of the signal in the sample to that of the cutoff value (termed the signal-to-cutoff ratio). Samples with low positive signal-to-cutoff ratios are usually negative on confirmatory tests and, thus, are usually false-positive results *(65)*. In 2003, the Centers for Disease Control and Prevention revised their earlier recommendations and advised that laboratories should routinely perform confirmatory tests on samples with low signal-to-cutoff ratio and report those with negative confirmatory test results as negative for anti-HCV *(66)*. The sensitivity of anti-HCV assays is thought to be in the range of 97–99% in detecting HCV infection; most false-negative results occur in the setting of immunosupression, such as with human immunodeficiency virus infection or in renal failure *(12,67)*. Antibodies are detectable at an average of 10–11 wk after infection with second-generation anti-HCV assays, but by 7–8 wk after infection with third-generation anti-HCV assays *(68)*. At the time of presentation with acute hepatitis resulting from HCV, over 40% of persons lack detectable anti-HCV *(69)*.

The major confirmatory test for anti-HCV involves isolation of HCV antigens on a solid support medium and determination of the number of antigens to which antibodies are present in serum. This supplemental testing is commonly referred to as recombinant immunoblot assay (RIBA), although a number of techniques are actually employed *(70)*. Samples with at least 1+ reactivity to two or more HCV antigens are termed RIBA positive and indicate a true positive antibody result. Samples with reactivity to only a single HCV antigen or with reactivity to two or more antigens plus the nonspecific yeast protein marker superoxide dismutase are termed RIBA indeterminate. Samples with less than 1+ reactivity to all antigens are termed RIBA negative and are considered false-positive results. Use of RIBA is most helpful in samples with low signal-to-cutoff ratio; samples with high signal-to-cutoff ratio are virtually never RIBA negative *(66)*. In acute hepatitis resulting from HCV, RIBA might be negative even though anti-HCV is positive *(66)*. With

Table 2
Pattern of Hepatitis C Laboratory Tests in Various Stages of Infection

Stage of Infection	*Anti-HCV*	*RIBA*	*HCV RNA*	*Core Antigen*
Early incubation (<2 wk)	Neg	Neg	Neg	Neg
Early acute infection (2 to 8–10 wk)	Neg	Neg	Pos	Pos
Acute infection	Pos/Neg[a]	Neg/Pos[b]	Pos/Neg[c]	Pos/Neg[c]
Chronic infection	Pos	Pos	Pos	Pos
Early viral clearance	Pos	Pos	Pos	Neg[d]
Spontaneous or treatment-induced clearance	Pos	Pos	Neg	Neg
False-positive result	Pos	Neg	Neg	Neg

Abbreviation: Pos = positive; Neg = negative.
[a]Anti-HCV is negative in 40–50% at time or presentation.
[b]RIBA might be negative even with positive anti-HCV during acute stage of infection.
[c]HCV RNA and core antigen might be transiently negative with development of antibody, even if chronic infection will develop.
[d]Core antigen is lost before HCV RNA; detection limit is equivalent to 20,000–50,000 IU/mL.

the third-generation RIBA, isolated antibodies to NS5 have never been associated with HCV viremia and are thought to represent a false-positive result *(65,71–73)*.

A more recent serologic test detects the presence of the HCV core antigen *(74)*. Because much of core antigen is attached to HCV antibodies, an initial step to free the antigen from binding antibodies precedes the quantitative sandwich immunoassay procedure *(75)*. Several studies have shown that the assay results agree reasonably well with HCV viral load, although the correlation coefficient is generally only about 0.8 *(76–78)*. HCV core antigen and HCV RNA rise and fall at similar times with acute infection *(79)* and fall in parallel during treatment of HCV *(76)*. The detection limit of the assay seems to be higher than that of HCV RNA assays, in the neighborhood of 20,000 to 50,000 international units HCV RNA per milliliter *(76)*. In untreated persons with HCV infection, the clinical sensitivity and specificity of HCV core antigen are similar to those for HCV RNA.

A summary of serologic and virologic tests in different phases of HCV infection is given in Table 2.

3. HEPATITIS C RNA MOLECULAR TESTS

3.1. HEPATITIS C RNA At present, the major method to determine the presence of active HCV infection is HCV RNA measurement. With acute HCV infection, HCV RNA becomes detectable by about 2–4 wk after infection, and viral loads climb rapidly *(36)*. With onset of acute hepatitis and development of antibodies to HCV, HCV RNA falls and might become undetectable, even in individuals who will progress to become chronically infected *(80)*. Over the first months to years of infection, HCV RNA might be transiently negative, but gradually increases by an average 0.3 log per year *(81,82)*. Average HCV viral loads are approx 2–3 million copies per milliliter. The dynamic range of various commercial HCV RNA assays is illustrated in Fig. 2. Once HCV infection is established beyond the first 6 mo, there is virtually no likelihood that the virus will clear spontaneously, although viremia might diminish in the late stages of decompensated cirrhosis *(83)*.

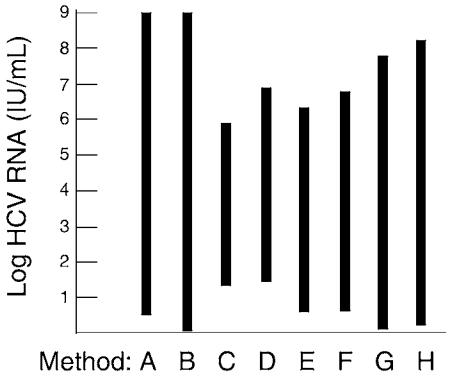


Fig. 2. Dynamic range of commercial HCV RNA assays. The bars represent the relative range of detection of HCV RNA assays that are currently available commercially. A and B: qualitative assays; C–H: quantitative assays. A: Roche Amplicor; B: Bayer TMA; C: Roche Amplicor Monitor 2.0; D: Bayer Versant 3.0; E: NGI SuperQuant; F: Roche Real Time PCR; G: Quest Heptimax; H: LabCorp QuantaSure.

Sample handling is important for HCV RNA measurements. Hepatitis C RNA is degraded by enzymes found in white blood cells. The activity of these enzymes is inhibited by calcium chelating agents, such as EDTA; many laboratories recommend collection of samples in EDTA plasma. It is critical to rapidly separate serum or plasma from cells by centrifugation, ideally within 1 h of collection, although EDTA plasma is stable in contact with cells at room temperature up to 24 h after collection *(84)*. Even if serum is not physically separated after centrifugation, samples show no significant loss of HCV RNA for up to 6 h. Serum or plasma physically separated from cells is stable at room temperature for up to 3 d, at refrigerator temperatures for up to 1 wk, and frozen at –70° C for years *(85,86)*. Although manufacturers often suggest use of expensive plasma separation gel systems, ordinary serum separator gel tubes appear adequate, as long as centrifugation is done soon after collection *(86,87)*. Heparin should never be used as an anticoagulant, as it inhibits molecular amplification methods.

3.2. QUALITATIVE HCV RNA Qualitative assays determine only the presence or absence of HCV RNA. Two primary methods are used in such assays: reverse transcriptase–polymerase chain reaction (RT-PCR) and transcription-mediated amplification (TMA). Many laboratories have established their own home-brew methods for qualitative HCV RNA determination using RT-PCR and primers from conserved regions of the HCV genome. The only US Food and Drug Administration (FDA)-approved molecular tests for HCV RNA involve assays developed by Roche Molecular Diagnostics, the Amplicor and Cobas Amplicor assays, both utilizing RT-PCR *(87)*. The detection limit of these assays is 50 IU/mL. The principles of Rt-PCR are the same as those described elsewhere (Chapter 5), so that details are not provided here. The major concern of using qualitative PCR methods in microtiter plates, as with all other targets of amplification, is the possibility of "splash-over" from one well to another.

The other major approach to qualitative testing is TMA, marketed by Bayer Diagnostics but not available to clinical

laboratories in kit form. In TMA, viral RNA hybridizes with oligonucleotides (containing a T7 promoter) bound to magnetic particles. After washing, incubation with reverse transcriptase and a second primer produces DNA containing the promoter sequence. A T7 RNA polymerase allows production of new RNA templates and amplification only of T7 labeled sequences, specifically increasing the amount of the original RNA exponentially. As with PCR, splash from well to well can occur. The detection limit of TMA (5 IU/mL) is approx 1 log lower than that of the Amplicor PCR. This difference could be clinically important in monitoring individuals who have been treated for HCV infection; several studies have shown that some individuals with undetectable HCV RNA using the Amplicor have detectable HCV RNA using TMA *(88,89)*. In contrast, there is a high degree of concordance between results of the two assays in initial evaluation of individuals found to be anti-HCV positive *(90,91)*. This is related to the high viral loads seen in untreated individuals; in data from 602 consecutive untreated HCV RNA positive individuals, we found that the average viral load was 1.4×10^6 international units, and only 11 individuals (2%) had a viral load $<10^4$ IU/mL (Dufour, unpublished observations).

3.3. QUANTIATIVE HCV RNA The methods used for qualitative HCV RNA determination can be adapted to quantitative measurement by using a known amount of a synthetic standard and comparing the amounts of HCV amplified to the amount of standard amplified using a calibration curve. To date, only one quantitative HCV RNA assay has been approved by the US FDA: the Bayer Versant™ HCV RNA 3.0 Assay. Determination of HCV viral load has become standard of care in evaluating patients before and during treatment. Until a few years ago, quantification was expressed in copies or "genome-equivalents" per volume of sample. This led to marked discrepancies in results between methods. The World Health Organization developed an international standard for HCV quantification using a purfied strain of genotype 1b *(92)*. Since then, commercial manufacturers have begun using the standard to calibrate their methods, and report results in arbitrary international units. Although this has standardized the reporting of viral load to a significant degree *(93)*, there are still some issues in comparability of methods particularly in unusual host situations (such as HIV coinfected individuals) *(94)*.

The two most commonly used methods for quantifying HCV RNA are commercially available kit assays: the Roche Amplicor Monitor 2.0 PCR *(95)* (and its semiautomated Cobas version) and the Bayer Versant™ 3.0 branched DNA (bDNA) method *(96)*. The relative dynamic ranges of these assays, along with those of other qualitative and quantitative assays, are shown in Fig. 2. The bDNA method uses a signal, rather than target, amplification method. After HCV RNA attaches to a target probe and the microtiter plate is washed, an extender probe to HCV RNA is added. A "preamplifier" sequence that hybridizes to the opposite end of the extender probe is then added to enhance sensitivity, followed by addition of an amplifier that contains multiple sites for attachment of oligonucleotides bound to alkaline phosphatase. These two assays have similar lower detection limits, but the bDNA method adds approx 1 log to the upper detection limit *(97)*. Although about 70% of untreated HCV RNA positive individuals have viral loads above the upper detection limit of the PCR assay *(98)*, only 2.5% of 602 consecutive HCV RNA positive samples were above the upper detection limit of the bDNA assay (Dufour, unpublished observations).

A recent modification of PCR, termed real-time PCR, is currently available as an analyte-specific reagent, allowing reduced carryover amplification, more rapid detection of amplification, increased low-end sensitivity, and a wider dynamic range *(99)*. Most commonly, this is accomplished by attaching probes with two labeling sequences attached. In the case of the Roche TaqMan system, one is a flurophore and the other a quencher; sequential amplification destroys the quencher and allows detection of fluorescence in direct proportion to the amount of DNA produced. With fluorescence resonance energy transfer, two oligonucleotide probes directed against adjacent sequences in the DNA are labeled with fluorescent dyes; amplification of DNA leads to excitation of one probe by energy released from the other, with the amount of fluorescence in direct proportion to amplified DNA. Fluorescent SYBR Green I dye, which binds nonspecifically to DNA, increases its fluorescence when bound. Use of standards with each type of procedure allows quantification of the amount of RNA in the original sample *(98)*.

In the United States, two large commercial laboratories have adapted one or more of these technologies to offer integrated tests that have a very wide dynamic range. Quest Diagnostics Heptimax combines a quantitative TMA assay with real-time PCR, whereas LabCorp's Quantasure utilizes real-time PCR. These assays provide similar sensitivity to qualitative assays, but upper measurable limits that minimize or eliminate the need for dilution of samples to produce a result.

3.4. HEPATITIS C GENOTYPE Several techniques have been developed to determine the particular genotype and subtype of HCV causing infection in an individual. As discussed earlier, such techniques typically target the 5′-untranslated and/or core regions of the HCV genome, the most highly conserved regions. Because most amplification methods for HCV RNA also target the 5′-untranslated region, qualitative PCR methods can be used to provide amplified RNA for use in determining the genotype.

Probably the most widely used technique is a commercial line probe assay *(100)*. A large number of oligonucleotide sequences are immobilized on a membrane, and incubated with amplified RNA; incubation with a color reagent detects areas of hybridization. The line probe assay allows recognition of most HCV types and subtypes accurately, although there are several subtypes that cannot be distinguished from one another. Overall concordance with other techniques is high *(101,102)*.

Direct sequencing of specific sequences of the HCV genome is felt to be the most accurate method available. Several commercial assays are available for sequence analysis of the 5′-untranslated region *(90)*, one using the NS5b region is in development. Although highly accurate, a small fraction of isolates cannot be identified because of variants in these highly conserved regions.

Serotyping assays, detecting antibodies to specific genotypes of HCV RNA, have been available for a number of years. Although they have improved in their accuracy, serotype assays

are accurate in only about 90% of cases *(103,104)* and sometimes detect antibody to more than one strain *(105)*, making it impossible to tell which strain currently infects the individual. Guidelines for hepatitis testing developed by the National Academy of Clinical Biochemistry and the American Association for the Study of Liver Diseases do not recommend use of serotyping assays for determination of HCV genotype *(106)*.

3.5. OTHER HCV MOLECULAR TESTS A number of investigators have evaluated the ability of sequence variability in the NS5a region to predict response or nonresponse to interferon. Because this area codes for an interferon-response element, it is reasonable to assume that this would be useful in determining which patients should not be treated with interferon. Although a number of studies have shown an inverse correlation between the number of mutations in this region and rate of response to treatment *(107–109)*, others have not *(110,111)* or have shown correlations only in some genotypes *(112)*. The number of patients evaluated in such studies is small, perhaps causing lack of detection of a correlation even if one exists. Another possible factor is that most of the studies showing a response were from Japan; Nakano found that there are differences in the sequences found in isolates from Japan and in other parts of the world and that a correlation between interferon response and NS5a mutations was seen only for the Japanese strain, even among Japanese individuals *(113)*.

4. CLINICAL USE OF HEPATITIS C TESTS

A variety of tests are used for diagnosis and monitoring of HCV infection. Tests used differ during different stages of disease and are best categorized by clinical category

4.1. ACUTE HEPATITIS C Although all individuals with HCV go through an incident of initial infection, in most persons it is clinically silent. Acute infection with HCV is most likely to be detected when it occurs following a needle-stick exposure from a person with known HCV or when it produces symptomatic infection and jaundice (estimated to occur in less than one-third of cases) *(37,38)*. There is evidence that those who develop jaundice are actually more likely to clear the infection and not progress to chronic hepatitis *(41)*.

During an incubation period that on average is 2 wk, HCV RNA is either undetectable or only intermittently detected. Following this, there is a period of rapid increase in the amount of circulating virus, with an estimated doubling time of less than 24 h *(36)*. Viral loads reach very high levels during this time, typically reaching values of 10^7 IU/mL and occasionally higher *(36)*. After an additional 1–2 mo, there is evidence of liver injury (characterized most reliably by increases in serum activities of liver enzymes alanine aminotransferase [ALT] and aspartate aminotransferase [AST]). Approximately 40–50% of persons with acute HCV infection that are detected are found during this stage, before development of anti-HCV *(69)*. After an average of 7–8 wk (with third-generation immunoassays) or 10–12 wk (second-generation immunoassays), anti-HCV becomes detectable *(68)*. At the time of this seroconversion, viral load decreases, sometimes to undetectable levels. In most individuals who will progress to chronic hepatitis and in some who will clear the infection, it remains detectable, but at lower levels. Recombinant immunoblot assay (RIBA) is often negative during the period of acute hepatitis and should not be used to classify an anti-HCV result as a false positive in the setting *(66)*.

In a person suspected of having acute HCV infection, the most reliable test for proving exposure is HCV RNA. Because of the high viral loads seen, either qualitative or quantitative assays would be acceptable for this purpose. Detectable HCV RNA in the absence of anti-HCV is strong evidence of recent HCV infection *(114)*. If an individual has low-titer anti-HCV and is HCV RNA positive, it is wise to repeat the anti-HCV at a later date; a high-titer antibody on subsequent testing supports a diagnosis of acute HCV infection. If an individual has a clinical picture of acute hepatitis and has undetectable HCV RNA but high positive anti-HCV, it is not possible to determine whether this represents recent or past exposure to HCV. If ALT remains elevated and HCV RNA becomes detectable on follow-up testing, this also supports a diagnosis of acute HCV infection.

4.2. CHRONIC HEPATITIS C Most individuals with chronic HCV infection are either asymptomatic or have only nonspecific symptoms. Detection of chronic HCV infection typically occurs when a person is found to have chronically increased ALT and is suspected to have chronic viral infection or is screened for viral infection because of risk factors, blood donation, or as part of an insurance or pre-employment physical examination. Such testing typically starts with anti-HCV. As mentioned earlier, CDC guidelines recommend that laboratories classify results as high positive or low positive based on the ratio of the signal in a sample to that of the cutoff between positive and negative results *(66)*. Samples with low signal : cutoff ratios are likely to be false positive *(65)* and should be confirmed by RIBA before being reported as positive. Routine confirmatory testing is not recommended if the signal : cutoff ratio is high, because very few, if any, of these will be RIBA negative *(65,66)*.

A truly positive anti-HCV could reflect a past infection with HCV or current infection. In persons who have cleared HCV RNA, antibody titers fall with time after infection and might eventually become undetectable in up to one-third of such individuals after 20 yr; however, most remain high positive *(115–117)*. For this reason, it is necessary to determine whether persons with positive anti-HCV are currently infected by measurement of HCV RNA. Most authorities recommend qualitative HCV RNA determinations for this purpose *(8,118)*. As indicated earlier, most infected individuals have viral loads well above the detection limits of quantitative assays. Use of quantitative HCV RNA is a better choice if the person would be a treatment candidate, as viral load should be determined before treatment (see below); because quantitative assays with similar detection limits to qualitative tests are developed, it is likely that quantitative assays will replace qualitative assays *(98)*. An undetectable HCV RNA suggests that infection has been cleared; however, a small number of persons with chronic infection might have intermittent viremia *(83)* or have a falsely negative HCV RNA resulting from sample-handling problems. In such cases, ALT measurements can provide a useful clue as to the true RNA status of the patient; most individuals with positive anti-HCV and normal ALT are RNA negative, whereas 95% of those with positive anti-HCV and multiple elevated ALT

Table 3
Laboratory Tests of Use in Evaluation of Treatment of Chronic Hepatitis C and Their Interpretation

PreTreatment
Quantitative HCV RNA (diluted, if needed, to report final result)
HCV genotype
After 12 wk of treatment
Early virologic response (EVR): Quantitative HCV RNA or core antigen undetectable or >2 log fall in HCV RNA or core antigen—continue treatment
Detectable and <2 log fall in HCV RNA or core antigen—consider discontinuation of treatment
After 24 wks of treatment
HCV RNA (using assay with lower detection limit of 50 IU/mL or lower): undetectable—continue treatment if genotype not 2 or 3; treatment finished if genotype 2 or 3; detectable—treatment failure; discontinue treatment
At end of treatment (24 wks for genotype 2 or 3, 48 wk for other genotypes)
End of Treatment (EOT) HCV RNA (using assay with lower detection limit of 50 IU/mL or lower): undetectable—on-treatment responder; repeat at 24 wks off therapy detectable—treatment failure
24 wk posttreatment HCV RNA (using assay with lower detection limit of 50 IU/mL or lower) undetectable—sustained virologic response (SVR): Detectable—virologic relapse

activity values are HCV RNA positive *(65)*. For those with elevated ALT, repeat HCV RNA determination is advisable.

4.3. TESTING RELATED TO TREATMENT OF HEPATITIS C

In contrast to other chronic viral infections such as hepatitis B or HIV, treatment has been successful in eradicating replicating HCV and halting progression of liver damage. Laboratory tests are important in determining the proper approach to treatment, and to determining its effectiveness (Table 3).

Interferon -α_2 is the only agent currently approved for treatment of chronic HCV. There are currently two potential approaches to treatment of chronic HCV: interferon alone, or a combination of interferon plus ribavirin. Although ribavirin is ineffective as a single agent for treating HCV *(119)*, it increases the effectiveness of interferon, increasing the number of responders by two or three times *(120)*. For many years, the only form of interferon available was "standard dose" interferon. Large doses (typically 3 million units) were injected three times a week. The short half-life of interferon led to widely fluctuating interferon levels. In 2001, the first longer-acting form of interferon, modified by attachment of polyethylene glycol (pegylated interferon), was approved for use in treating HCV. Use of pegylated interferon results in sustained, high interferon levels, requiring injection only once weekly, and improved response rates. Currently, the preferred treatment for chronic HCV infection is the combination of pegylated interferon plus ribavirin *(121)*.

Duration of treatment is dependent on the genotype of HCV. Genotypes 2 and 3 respond much better to treatment; only 24 wk of therapy is needed to achieve maximum benefit, compared to 48 wk in persons infected with other genotypes *(121)*. A number of other factor influence likelihood of response, as illustrated in Table 4. Currently, treatment is considered in all patients except those with decompensated cirrhosis, where treatment might lead to worsening of the patient's condition *(122,123)*.

Table 4
Factors Associated with Increased Likelihood of Response to Antiviral Therapy of Hepatitis C

Female gender
Age < 40 yr
Moderate or lower fibrosis on biopsy
Caucasian race
Viral load < median (800,000 IU/mL)
Genotype 2 or 3
Low body mass index

4.4. PRETREATMENT EVALUATION

When a decision is made to treat a person with chronic HCV, it is important to determine both viral load and genotype. Although individuals with high viral load have a reduced rate of response to treatment, the main reason to determine viral load is for early determination of response to treatment (see below). Because viral load increases an average of 0.25 log per year *(82)*, it is not critical that viral load be repeated if a previous viral load was determined in the past 2 yr. As mentioned earlier, genotype is important both as a predictor of treatment response and as a determinant of duration of therapy. It is essential to determine genotype before therapy is started, because those who respond to treatment will not have detectable virus, so that genotype determination will be impossible.

4.5. MONITORING TREATMENT EFFECTIVENESS

Once treatment is initiated, the most reliable means to determine efficacy is to evaluate the response of HCV RNA. Successful treatment is associated with at least two different phases of viral clearance *(124)*. The first phase, which occurs rapidly over the course of days, is thought to reflect HCV RNA clearance from a circulating pool through the antiviral effect of interferon. In the second phase of clearance, infected liver cells (the major site of viral replication) undergo cell turnover and are replaced by uninfected cells. The second phase of clearance is more variable in duration. First-phase clearance is less specific for detecting success of antiviral treatment; therefore, it is necessary to evaluate whether second-phase clearance has occurred.

At present, the most reliable early sampling time is felt to be after 12 wk of treatment. The term "early virologic response" (EVR) refers to a fall of HCV RNA to undetectable, or at least by 2 logs, from pretreatment to 12 wk on treatment *(125)*. Individuals who do not have an EVR have less than 2% likelihood of successful treatment and should be recommended to discontinue treatment *(125)*. In those with an EVR, treatment should be continued to 24 wk, and HCV RNA should be rechecked using an assay with a lower detection limit of 50 IU/mL or lower *(126)*. If virus remains detectable, the likelihood of viral clearance is less than 1%, and discontinuation of treatment is usually recommended. [There are some data that treatment with interferon may lower risk of progression to cirrhosis *(127)* and lower risk of hepatocellular carcinoma *(128)*, even in those who do not clear the virus; at present, no guidelines recommend continuation of treatment for this purpose, although controlled trials of this form of therapy are ongoing.]

In those who have evidence of viral clearance, viral RNA should be rechecked using an assay with a detection limit of

50 IU/mL or lower at the end of treatment (EOT) and again 6 mo following completion of treatment. Undetectable virus at the 6-mo follow-up is termed a sustained virologic response (SVR). Individuals who achieve SVR have over 99% likelihood of remaining free of virus in long-term follow-up *(126)*; for all practical purposes, they can be considered to have cleared the virus.

5. BLOOD DONOR TESTING

The other major setting for use of HCV RNA testing is blood donors. Although antibody screening is highly effective in preventing transmission of HCV, even third-generation immunoassays are associated with one case of posttransfusion HCV transmission for every 250,000 to 500,000 units transfused *(129)*. Many donor centers have now instituted screening using nucleic acid tests, either polymerase chain reaction or transcription-mediated amplification. Testing is typically performed on small pools currently *(130)*, but it is anticipated that reduced cost will allow screening of individual units. Use of testing on small pools (16–24 units per pool) has reduced the frequency of posttransfusion HCV infection to 1 case for every 1.6 million units transfused *(131)*.

REFERENCES

1. Alter, M., Kruszon-Moran, D., Nainan, O., et al. The prevalence of hepatitis C virus infection in the United States, 1988 through 1994. *N. Engl. J. Med.* **341:**556–562, 1999.
2. Kim, W. The burden of hepatitis C in the United States. *Hepatology* **36:**S30–S34, 2002.
3. Ho, S., Gebhard, R., Tetrick, L., et al. Key components and logistics of hepatitis C care within a VISN: a proposal. *Veterans Health Syst. J.* 41–45, 1999.
4. Brau, N., Bini, E., Shahidi, A., et al. Prevalence of hepatitis C and coinfection with HIV among United States veterans in the New York City metropolitan area. *Am. J. Gastroenterol.* **97:**2071–2078, 2002.
5. Mansell, C. and Locarnini, S. Epidemiology of hepatitis C in the East. *Semin. Liver Dis.* **15:**15–32, 1995.
6. Wasley, A. and Alter, M. Epidemiology of hepatitis C: geographic differences and temporal trends. *Semin. Liver Dis.* **20:**1–16, 2000.
7. Alter, M. Epidemiology of hepatitis C. *Hepatology* **26:**62S–65S, 1997.
8. Centers for Disease Control and Prevention. Recommendations for the prevention and control of hepatitis C virus (HCV) infection and HCV-related chronic disease. *MMWR* **47:** 1–39, 1998.
9. McCarthy, J. and Flynn, N. Hepatitis C in methadone maintenance patients: prevalence and public policy implications. *J. Addict. Dis.* **20:**19–31, 2001.
10. Miller, C., Johnston, C., Spittal, P., et al. Opportunities for prevention: hepatitis C prevalence and incidence in a cohort of young injection drug users. *Hepatology* **36:**737–742, 2002.
11. Murphy, E., Bryzman, S., Glynn, S., et al. Risk factors for hepatitis C virus infection in United States blood donors. *Hepatology* **31:** 756–762, 2002.
12. Bukh, J., Wantzin, P., Krogsgaard, K., Knudsen, F., Purcell, R., and Miller, R. High prevalence of hepatitis C virus (HCV) RNA in dialysis patients: failure of commercially available antibody tests to identify a significant number of patients with HCV infection. Copenhagen Dialysis HCV Study Group. *J. Infect. Dis.* **168:** 1343–1348, 1993.
13. Courouce, A., Le Marrec, N., Girault, A., Ducamp, S., and Simon, N. Anti-hepatitis C virus (anti-HCV) seroconversion in patients undergoing hemodialysis: comparison of second- and third-generation anti-HCV assays. *Transfusion* **34:**790-795, 1994.
14. Mitsui, T., Iwano, K., Masuko, K., et al. Hepatitis C virus infection in medical personnel after needlestick accident. *Hepatology* **16:** 1109–1114, 1992.
15. Okamoto, M., Nagata, I., Murakami, J., et al. Prospective reevaluation of risk factors in mother-to-child transmission of hepatitis C virus: high virus load, vaginal delivery, and negative anti-NS4 antibody. *J. Infect. Dis.* **182:**1511–1514, 2000.
16. Mazza, C., Ravaggi, A., Rodella, A., et al. Prospective study of mother-to-infant transmission of hepatitis C virus (HCV) infection. Study Group for Vertical Transmission. *J. Med. Virol.* **54:**12–19, 1998.
17. Roberts, E. and Yeung, L. Maternal-infant transmission of hepatitis C virus infection. *Hepatology* **36:**S106–S112, 2002.
18. Kao, J., Hwang, Y., Chen, P., et al. Transmission of hepatitis C virus between spouses: the important role of exposure duration. *Am. J. Gastroenterol.* **91:**2087–2090, 1996.
19. Osella, A., Massa, M., Joekes, S., et al. Hepatitis B and C virus sexual transmission among homosexual men. *Am. J. Gastroenterol.* **93:** 49–52, 1998.
20. Terrault, N. Sexual activity as a risk factor for hepatitis C. *Hepatology* **36:**S99–S105, 2002.
21. Wejstal, R. Sexual transmission of hepatitis C virus. *J. Hepatol.* **31 (Suppl. 1):**92–95, 1995.
22. Bresters, D., Mauser-Bunschoten, E., Reesink, et al. Sexual transmission of hepatitis C virus. *Lancet* **342:**210–211, 1993.
23. Conry-Cantilena, C., Van Raden, M., Gibble, J., et al. Routes of infection, viremia, and liver disease in blood donors found to have hepatitis C virus infection. *N. Engl. J. Med.* **334:**1691–1696, 1996.
24. Choo, Q.-L., Kuo, G., Weiner, A., Bradley, D., and Houghton, M. Isolation of a c-DNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* **244:**359–362, 1989.
25. Houghton, M., Weiner, A., Han, J., Kuo, G., and Choo, Q. Molecular biology of the hepatitis C viruses: implications for diagnosis, development, and control of viral disease. *Hepatology* **14:**381–388, 1991.
26. Honda, M., Kaneko, S., Sakai, A., Unoura, M., Murakami, S., and Kobayashi, K. Degree of diversity of hepatitis C virus quasispecies and progression of liver disease. *Hepatology* **20:**1144–1151, 1994.
27. Baumert, T., Wellnitz, S., Aono, S., et al. Antibodies against hepatitis C virus-like particles and viral clearance in acute and chronic hepatitis C. *Hepatology* **32:**610–617, 2000.
28. Rehermann, B. Interaction between the hepatitis C virus and the immune system. *Semin. Liver Dis.* **20:**127–144, 2000.
29. Frangeul, L., Cresta, P., Perrin, M., et al. Mutations in NS5A region of hepatitis C virus genome correlate with presence of NS5A antibodies and response to interferon therapy for most common European hepatitis C virus genotypes. *Hepatology* **28:**1674–1679, 1998.
30. Ideo, G. and Bellobuono, A. New therapies for the treatment of chronic hepatitis C. *Curr. Pharm. Des.* **8:**959–966, 2002.
31. Bukh, J., Miller, R., and Purcell, R. Genetic heterogeneity of hepatitis C virus: quasispecies and genotypes. *Semin. Liver. Dis.* **15:**41–63, 1995.
32. Farci, P., Shimoda, A., Coiana, A., et al. The outcome of acute hepatitis C predicted by the evolution of the viral quasispecies. *Science* **288:**339–344, 2000.
33. Lin, H. J., Seeff, L. B., Barbosa, L., and Hollinger, F. B. Occurrence of identical hypervariable region 1 sequences of hepatitis C virus in transfusion recipients and their respective blood donors: divergence over time. *Hepatology* **34:**424–429, 2001.
34. Gao, G., Buskell, Z., Seeff, L., and Tabor, E. Drift in the hypervariable region of the hepatitis C virus during 27 years in two patients. *J. Med. Virol.* **68:**60–67, 2002.
35. El-Awady, M., Ismail, S., El-Sagheer, M., Sabour, Y., Amr, K., and Zaki, E. Assay for hepatitis C virus in peripheral blood mononuclear cells enhances sensitivity of diagnosis and monitoring of HCV-associated hepatitis. *Clin. Chim. Acta.* **283:**1–14, 1999.
36. Naito, M., Hayashi, N., Hagiwara, H., et al. Serial quantitative analysis of serum hepatitis C virus RNA level in patients with acute and chronic hepatitis C. *J. Hepatol.* **20:**755–759, 1994.
37. Seeff, L., Wright, E., Zimmerman, H., and McCollum, R. W. VA cooperative study of post-transfusion hepatitis, 1969–1974: incidence and characteristics of hepatitis and responsible risk factors. *Am. J. Med. Sci.* **270:**355–362, 1975.
38. Healey, C., Sabharwal, N., Daub, J., et al. Outbreak of acute hepatitis C following the use of anti-hepatitis C virus—screened intravenous immunoglobulin therapy. *Gastroenterology* **110:**1120–1126, 1996.

39. Hino, K., Sainokami, S., Shimoda, K., Niwa, H., and Iino, S. Clinical course of acute hepatitis C and changes in HCV markers. *Dig. Dis. Sci.* **39:**19–27, 1994.
40. Giuberti, T., Marin, M., Ferrari, C., et al. Hepatitis C virus viremia following clinical resolution of acute hepatitis C. *J. Hepatol.* **20:** 666–671, 1994.
41. Villano, S., Vlahov, D., Nelson, K., Cohn, S., and Thomas, D. Persistence of viremia and the importance of long-term follow-up after acute hepatitis C infection. *Hepatology* **29:**908–914, 1999.
42. Datz, C., Cramp, M., Haas, T., et al. The natural course of hepatitis C virus infection 18 years after an epidemic outbreak of non-A, non-B hepatitis in a plasmapheresis centre. *Gut.* **44:**563–567, 1999.
43. Alter, H. and Seeff, L. Recovery, persistence, and sequelae in hepatitis C virus infection: a perspective on long-term outcome. *Semin. Liver Dis.* **20:**17–35, 2000.
44. Gholson, C., Morgan, K., Catinis, G., et al. Chronic hepatitis C with normal aminotransferase levels: a clinical histologic study. *Am. J. Gastroenterol.* **92:**1788–1792, 1997.
45. Dincer, D., Okten, A., Kaymakoglu, S., et al. Persistently normal alanine transaminase levels in chronic C hepatitis: what does it tell us? *Hepatogastroenterology* **48:**1397–1400, 2001.
46. Pradat, P., Alberti, A., Poynard, T., et al. Predictive value of ALT levels for histologic findings in chronic hepatitis C: a European collaborative study. *Hepatology* **36:**973–977, 2002.
47. McCormick, S., Goodman, Z., Maydonovitch, C., and Sjogren, M. Evaluation of liver histology, ALT elevation, and HCV RNA titer in patients with chronic hepatitis C. *Am. J. Gastroenterol.* **91:**1516–1522, 1996.
48. Dufour, M., Stinson, F., and Caces, M. Trends in cirrhosis morbidity and mortality: United States, 1979-1988. *Semin. Liver Dis.* **13:**109–125, 1993.
49. Alberti, A., Chemello, L., and Benvegnu, L. Natural history of hepatitis C. *J. Hepatol.* **31(Suppl. 1):**17–24, 1999.
50. Seeff, L. Natural history of chronic hepatitis C. *Hepatology* **36:**S35–S46, 2002.
51. Vogt, M., Lang, T., Frosner, G., et al. Prevalence and clinical outcome of hepatitis C infection in children who underwent cardiac surgery before the implementation of blood-donor screening. *N. Engl. J. Med.* **341:**866–870, 1999.
52. Kenny-Walsh, E. Clinical outcomes after hepatitis C infection from contaminated anti-D immune globulin. *N. Engl. J. Med.* **340:**1228–1233, 1999.
53. Wiese, M., Berr, F., Lafrenz, M., Porst, H., and Oesen, U. Low frequency of cirrhosis in a hepatitis C (genotype 1b) single-source outbreak in Germany: a 20-year multicenter study. *Hepatology* **32:** 91–96, 2000.
54. Fattovich, G., Giustina, G., Degos, F., et al. Morbidity and mortality in compensated cirrhosis type C: a restrospective follow-up study of 384 patients. *Gastroenterology* **112:**463–472, 1997.
55. El-Serag, H. and Mason, A. Rising incidence of hepatocellular carcinoma in the United States. *N. Engl. J. Med.* **340:**745–750,1999.
56. Davis, G., Albright, J., Cook, S., et al. Projecting the future healthcare burden from hepatitis C in the United States. *Hepatology* **28:**390a, 1998.
57. El-Serag, H. Epidemiology of hepatocellular carcinoma. *Clin. Liver Dis.* **5:**87–107, 2001.
58. Yang, H., Lu, S., Liaw, Y., et al. Hepatitis B e antigen and the risk of hepatocellular carcinoma. *N. Engl. J. Med.* **347:**168–174, 2002.
59. Di Bisceglie, A. Hepatitis C and hepatocellular carcinoma. *Hepatology* **26:**34S–38S, 1997.
60. Fattovich, G. Progression of hepatitis B and C to hepatocellular carcinoma in Western countries. *Hepatogastroenterology* **45(Suppl. 3):** 1206–1213, 1998.
61. El-Serag, H., Hampel, H., Yeh, C., and Rabeneck, L. Extrahepatic manifestations of hepatitis C among United States male veterans. *Hepatology* **36:**1439–1445, 2002.
62. Kayali, Z., Buckwold, V., Zimmerman, B., and Schmidt, W. Hepatitis C, cryoglobulinemia, and cirrhosis: a meta-analysis. *Hepatology* **36:**978–985, 2002.
63. Hermine, O., Lefrere, F., Bronowicki, J., et al. Regression of splenic lymphoma with villous lymphocytes after treatment of hepatitis C virus infection. *N. Engl. J. Med.* **347:**89–94, 2002.
64. Mehta, S., Brancati, F., Sulkowski, M., Strathdee, S., Szklo, M., and Thomas, D. Prevalence of type 2 diabetes mellitus among persons with hepatitis C virus infection in the United States. *Ann. Intern. Med.* **133:**592–599, 2000.
65. Dufour, D., Talastas, M., Fernandez, M., Harris, B., Strader, D., and Seeff, L. Low positive anti-hepatitis C virus enzyme immunoassay results: an important predictor of low likelihood of hepatitis C infection. *Clin. Chem.* **49:**479–486, 2003.
66. Alter, M., Kuhnert, W., and Finelli, L. Guidelines for laboratory testing and results reporting of antibody to hepatitis C virus. *MMWR* **52:**1–15, 2003.
67. Marcellin, P., Martinot-Peignoux, M., Elias., et al. Hepatitis C virus (HCV) viremia in human immunodeficiency virus-seronegative and -seropositive patients with indeterminate HCV recombinant immunoblot assay. *J. Infect. Dis.* **170:**433–435, 1994.
68. Gretch, D. Diagnostic tests for hepatitis C. *Hepatology* **26:** 43S–47S, 1997.
69. Barrera, J., Bruguera, M., Ercilla, M., et al. Persistent hepatitis C viremia after acute self-limiting posttransfusion hepatitis C. *Hepatology* **21:**639–644, 1995.
70. Courouce, A., Noel, L., Barin, F., et al. A comparative evaluation of the sensitivity of five anti-hepatitis C virus immunoblot assays. *Vox Sang* **74:**217–224, 1998.
71. Damen, M., Zaaijer, H., Cuypers, H., et al. Reliability of the third-generation recombinant immunoblot assay for hepatitis C virus. *Transfusion* **35:**745–749, 1995.
72. Vernelen, K., Claeys, H., Verhaert, H., Volckaerts, A., and Vernylen, C. Significance of NS3 and NS5 antigens in screening for HCV antibody. *Lancet* **343:**853, 1994.
73. LaPerche, S., Courouce, A., Lemaire, J., Coste, J., Defer, C., and Cantaloube, J. GB virus type C/hepatitis G virus infection in French blood donors with anti-NS5 isolated reactivities by recombinant immunoblot assay for hepatitis C virus. *Transfusion* **39:**790–791, 1999.
74. Aoyagi, K., Ohue, C., Iida, K., et al. Development of a simple and highly sensitive enzyme immunoassay for hepatitis C virus core antigen. *J. Clin. Microbiol.* **37:**1802–1808, 1999.
75. Peterson, J., Green, G., Iida, K., et al. Detection of hepatitis C core antigen in the antibody negative 'window' phase of hepatitis C infection. *Vox Sang* **78:**80–85, 2000.
76. Bouvier-Alias, M., Patel, K., Dahari, H., et al. Clinical utility of total HCV core antigen quantification: a new indirect marker of HCV replication. *Hepatology* **36:**211–218, 2002.
77. Kurtz, J., Boxall, E., Qusir, N., Shirley, J., Coleman, D., and Chandler, C. The diagnostic significance of an assay for 'total' hepatitis C core antigen. *J. Virol. Methods.* **96:**127–132, 2001.
78. Tanaka, E., Ohue, C., Aoyagi, K., et al. Evaluation of a new enzyme immunoassay for hepatitis C virus (HCV) core antigen with clinical sensitivity approximating that of genomic amplification of HCV RNA. *Hepatology* **32:**388–393, 2000.
79. Kobayashi, M., Tanaka, E., Matsumoto, A., et al. Clinical application of hepatitis C virus core protein in early diagnosis of acute hepatitis C. *J. Gastroenterol.* **33:**508–511, 1998.
80. Tan, D., Im, S., Peng, W., and Ng, M. Follow-up study of acute hepatitis C. *Arch. Virol.* **138:**71-84, 1994.
81. Nguyen, T., Sedghi-Vaziri, A., Wilkes, L., et al. Fluctuations in viral load (HCV RNA) are relatively insignificant in untreated patients with chronic HCV infection. *J. Viral Hepat.* **3:**75–78, 1996.
82. Fanning, L., Kenny-Walsh, E., Levis, J., et al. Natural fluctuations of hepatitis C viral load in a homogeneous patient population: a prospective study. *Hepatology* **31:**225–229, 2000.
83. Yokosuka, O., Kojima, H., Imazeki, F., et al. Spontaneous negativation of serum hepatitis C virus RNA is a rare event in type C chronic liver diseases: analysis of HCV RNA in 320 patients who were followed for more than 3 years. *J. Hepatol.* **31:**394–399, 1999.
84. Davis, G., Lau, J., Urdea, M., et al. Quantitative detection of hepatitis C virus RNA with a solid-phase signal amplification method:

definition of optimal conditions for specimen collection and clinical application in interferon-treated patients. *Hepatology* **19:**1337–1341, 1994.
85. Grant, P., Kitchen, A., Barbara, J., et al. Effects of handling and storage of blood on the stability of hepatitis C virus RNA: implications for NAT testing in transfusion practice. *Vox Sang* **78:**137–142, 2000.
86. Trabaud, M., Bailly, F., Colucci, G., and Trepo, C. Stability of hepatitis C virus RNA in serum from samples collected in a closed-tube system for serum separation and transport, as measured by a quantitative competitive PCR assay. *J. Viral. Hepat.* **3:**207–209, 1996
87. Doglio, A., Laffont, C., Caroli-Bosc, F., Rochet, P., and Lefebvre, J. Second generation of the automated Cobas Amplicor HCV assay improves sensitivity of hepatitis C virus RNA detection and yields results that are more clinically relevant. *J. Clin. Microbiol.* **37:**1567–1569, 1999.
88. Comanor, L., Anderson, F., Ghany, M., et al. Transcription-mediated amplification is more sensitive than conventional PCR-based assays for detecting residual serum HCV RNA at end of treatment. *Am. J. Gastroenterol.* **96:**2968–2972, 2001.
89. Sarrazin, C., Teuber, G., Kokka, R., Rabenau, H., and Zeuzum, S. Detection of residual hepatitis C virus RNA by transcription-mediated amplification in patients with complete virologic response according to polymarase chain reaction-based assays. *Hepatology* **32:**818–823, 2000.
90. Ross, R., Viazov, S., Hoffmann, S., and Roggendorf, M. Performance characteristics of a transcription-mediated nucleic acid amplification assay for qualitative detection of hepatitis C virus RNA. *J. Clin. Lab. Anal.* **15:**308–313, 2001.
91. Krajden, M., Ziermann, R., Khan, A., et al. Qualitative detection of hepatitis C virus RNA: comparison of analytical sensitivity, clinical performance, and workflow of the Cobas Amplicor HCV Test Version 2.0 and the HCV RNA Transcription-Mediated Amplification Qualitative Assay. *J. Clin. Microbiol.* **40:**2903–2907, 2002.
92. Saldanha, J., Lelie, N., and Heath, A. Establishment of the first international standard for nucleic acid amplification technology (NAT) assays for HCV RNA. WHO Collaborative Study Group. *Vox Sang* **76:**149–158, 1999.
93. Ross, R., Viazov, S., Sarr, S., Hoffmann, S., Kramer, A., and Roggendorf, M. Quantitation of hepatitis C virus RNA by third generation branched DNA-based signal amplification assay. *J. Virol. Methods* **101:**159–168, 2002.
94. Sherman, K., Rouster, S., and Horn, P. Comparison of methodologies for quantification of hepatitis C virus (HCV) RNA in patients coinfected with HCV and human immunodeficiency virus. *Clin. Infect. Dis.* **35:**482–487, 2002.
95. Afonso, A., Didier, J., Plouvier, E., et al. Performance of an automated system for quantification of hepatitis C virus RNA. *J. Virol. Methods* **86:**55–60, 2000.
96. Trimoulet, P., Halfon, P., Pohier, E., Khiri, H., Chene, G., and Fleury, H. Evaluation of the VERSANT HCV RNA 3.0 assay for quantification of hepatitis C virus RNA in serum. *J. Clin. Microbiol.* **40:**2031–2036, 2002.
97. Germer, J., Heimgartner, P., Ilstrup, D., Harmsen, W., Jenkins, G., and Patel, R. Comparative evaluation of the VERSANT HCV RNA 3.0, QUANTIPLEX HCV RNA 2.0, and COBAS AMPLICOR HCV MONITOR Version 2.0 assays for quantification of hepatitis C virus RNA in serum. *J. Clin. Microbiol.* **40:**495–500, 2002.
98. Pawlotsky, J. Molecular diagnosis of viral hepatitis. *Gastroenterology* **122:**1554–1568, 2002.
99. Yang, J., Lai, J., Douglas, S., Metzger, D., Zhu, X., and Ho, W. Real-time RT-PCR for quantitation of hepatitis C virus RNA. *J. Virol. Methods* **102:**119–128, 2002.
100. van Doorn, L., Kleter, B., Stuyver, L., et al. Analysis of hepatitis C virus genotypes by a line probe assay and correlation with antibody profiles. *J. Hepatol.* **21:**122–129, 1994.
101. Stuyver, L., Wyseur, A., Arnhem, W. V., et al. Hepatitis C genotyping by means of 5'-UR/core line probe assays and molecular analysis of untypeable samples. *Virus Res.* **38:**137–157, 1995.
102. Mitchell, P., Sloan, L., Majewski, D., et al. Comparison of line probe assay and DNA sequencing of 5' untranslated region for genotyping hepatitis C virus: description of novel line probe patterns. *Diagn. Microbiol. Infect. Dis.* **42:**175–179, 2002.
103. Pawlotsky, J., Prescott, L., Simmonds, P., et al. Serological determination of hepatitis C virus genotype: comparison with a standardized genotyping assay. *J. Clin. Microbiol.* **35:**1734–1739, 1997.
104. Fabrizi, F., Martin, P., Quan, S., et al. Serotyping strip immunoblot assay for assessing hepatitis C virus strains in dialysis patients. *Am. J. Kidney Dis.* **35:**832–838, 2000.
105. Toyoda, H., Fukuda, Y., Hayakawa, T., et al. Presence of multiple genotype-specific antibodies in patients with persistent infection with hepatitis C virus (HCV) of a single genotype: evidence for transient or occult superinfection with HCV of different genotypes. *Am. J. Gastroenterol.* **94:**2230–2236, 1999.
106. Dufour, D., Lott, J., Nolte, F., Gretch, D., Koff, R., and Seeff, L. Diagnosis and monitoring of hepatic injury. I. Performance characteristics of laboratory tests. *Clin. Chem.* **46:**2027–2049, 2000.
107. Sakuma, I., Enomoto, N., Kurosaki, M., Izumi, N., Marumo, F., and Sato, C. Differential effect of interferon on hepatitis C virus 1b quasispecies in the nonstructural protein 5A gene. *J. Infect. Dis.* **180:**1001–1009, 1999.
108. Murakami, T., Enomoto, N., Kurosaki, M., Izumi, N., Marumo, F., and Sato, C. Mutations in nonstructural protein 5A gene and response to interferon in hepatitis C virus genotype 2 infection. *Hepatology* **30:**1045–1053, 1999.
109. Terazawa, Y., Yoshioka, K., Kobayashi, M., et al. Mutations in interferon sensitivity-determining region of hepatitis C virus: its relation to change in viral load. *Am. J. Gastroenterol* **95:**1781–1787, 2000.
110. Sarrazin, C., Berg, T., Lee, J., et al. Mutations in the protein kinase-binding domain of the NS5A protein in patients infected with hepatitis C virus type 1a are associated with treatment response. *J. Infect. Dis.* **181:**432–441, 2000.
111. Murphy, M., Rosen, H., Marousek, G., and Chou, S. Analysis of sequence configurations of the ISDR, PKR-binding domain, and V3 region as predictors of response to induction interferon-alpha and ribavirin therapy in chronic hepatitis C infection. *Dig. Dis. Sci.* **47:**1195–1205, 2002.
112. Saiz, J., Lopez-Labrador, F., Ampurdanes, S., et al. The prognostic relevance of the nonstructural 5A gene interferon sensitivity determining region is different in infections with genotype 1b and 3a isolates of hepatitis C virus. *J. Infect. Dis*. **177:**839–847, 1998.
113. Nakano, I., Fukuda, Y., Katano, Y., Nakano, S., Kumada, T., and Hayakawa, T. Why is the interferon sensitivity-determining region (ISDR) system useful in Japan? *J. Hepatol.* **30:**1014–1022, 1999.
114. Alberti, A., Boccato, S., Vario, A., and Benvegnu, L. Therapy of acute hepatitis C. *Hepatology* **36:**S195–S200, 2002.
115. Beld, M., Penning, M., Putten, M. V., et al. Quantitative antibody responses to structural (core) and nonstructural (NS3, NS4, and NS5) hepatitis C virus proteins among seroconverting injecting drug users: impact of epitope variation and relationship to detection of HCV RNA in blood. *Hepatology* **29:**1288–1298, 1999.
116. Rodger, A., Roberts, S., Lanigan, A., Bowden, S., Brown, T., and Crofts, N. Assessment of long-term outcomes of community-acquired hepatitis C infection in a cohort with sera stored from 1971 to 1975. *Hepatology* **32:**582–587, 2000.
117. Seeff, L., Hollinger, F., Alter, H., et al. Long-term mortality and morbidity of transfusion-associated non-A, non-B, and type C hepatitis: A National Heart, Lung, and Blood Institute collaborative study. *Hepatology* **33:**455–463, 2001.
118. Pawlotsky, J. Use and interpretation of virological tests for hepatitis C. *Hepatology* **36:**S65–S73, 2002.
119. Bodenheimer, H., Lindsay, K., Davis, G., Lewis, J., Thung, S., and Seeff, L. Tolerance and efficacy of oral ribavirin treatment of chronic hepatitis C: a multicenter trial. *Hepatology* **26:**473–477, 1997.
120. Cummings, K., Lee, S., West, E., et al. Interferon and ribavirin vs interferon alone in the re-treatment of chronic hepatitis C previously nonresponsive to interferon: A meta-analysis of randomized trials. *JAMA* **285:**193–199, 2001.

121. Di Bisceglie, A. and Hoofnagle, J. Optimal therapy of hepatitis C. *Hepatology* **36:**S121–S127, 2002.
122. Schalm, S., Fattovich, G., and Brouwer, J. Therapy of hepatitis C: patients with cirrhosis. *Hepatology* **26:**128S–132S, 1997.
123. Wright, T. Treatment of patients with hepatitis C and cirrhosis. *Hepatology* **36:**S185–S194, 2002.
124. Layden, T., Mika, B., and Wiley, T. Hepatitis C kinetics: mathematical modeling of viral response to therapy. *Semin. Liver Dis.* **20:**173–183, 2000.
125. Davis, G. Monitoring of viral levels during therapy of hepatitis C. *Hepatology* **36:**S145–S151, 2002.
126. Anon. National Institutes of Health Consensus Development Conference Statement: Management of hepatitis C: 2002—June 10–12, 2002. *Hepatology* **36:**S3–S20, 2002.
127. Dufour, J., De Lellis, R., and Kaplan, M. Regression of hepatic fibrosis in hepatitis C with long-term interferon treatment. *Dig. Dis. Sci.* **43:**2573–2576, 1998.
128. Bonino, F. Effect of interferon-α on progression of cirrhosis to hepatocellular carcinoma: A retrospective cohort study. *Lancet* **351:**1535–1539, 1998.
129. Galel, S., Strong, D., Tegtmeier, G., et al. Comparative yield of HCV RNA testing in blood donors screened by 2.0 versus 3.0 antibody assays. *Transfusion* **42:**1507–1513, 2002.
130. Hitzler, W. and Runkel, S. Screening of blood donations by hepatitis C virus polymerase chain reaction (HCV-PCR) improves safety of blood products by window period reduction. *Clin. Lab.* **47:**219–222, 2001.
131. Busch, M., Kleinman, S., and Nemo, G. Current and emerging infectious risks of blood transfusions. *JAMA* **289:**959–962, 2003.

38 Cytomegalovirus

ANGELA M. CALIENDO AND NELL S. LURAIN

1. INTRODUCTION

Cytomegalovirus (CMV) remains an important pathogen in immunocompromised individuals, including transplant recipients, persons with acquired immunodeficiency syndrome (AIDS), and those using corticosteroids or other immune-modulating drugs. The great majority of individuals are asymptomatic following primary infection, whereas a small number will develop a mononucleosis-type syndrome. After primary infection with CMV, a lifelong latent infection is established. In immunocompetent individuals, the virus will remain latent without causing any clinical symptoms. However, should an individual become immunocompromised, the virus can reactivate leading to active disease. Moreover, those who acquire their primary CMV infection while immunosuppressed are at risk for severe disease.

In persons with AIDS, CMV disease usually occurs when the CD_4 cell count is below 50–100 cells/mm^3, with retinitis, esophagitis, and colitis representing common clinical presentations. The occurrence and severity of CMV disease in transplant recipients is related to a number of factors, including the type of organ transplanted, the CMV serostatus of the organ recipient and donor, and the overall degree of immunosuppression. In general, CMVseronegative recipients of organs from CMV-seropositive donors are at risk for severe disease, because primary infection occurs during immunosuppression. Symptomatic disease can also occur in CMV-seropositive recipients regardless of the CMV serostatus of the donor. In these situations, CMV disease can result from reactivation of the recipient's endogenous strain and/or the donor strain. The clinical presentation of CMV disease in organ transplant recipients is diverse, including fever and leukopenia, interstitial pneumonitis, esophagitis, colitis, retinitis, and encephalitis.

The diagnostic challenge with CMV disease is to distinguish latent or low-level persistent infection from clinically significant reactivation. Infectious virus, viral antigen, and viral nucleic acid can all be detected in patients without clinically important disease. This is particularly a problem with very sensitive molecular assays that can detect as few as 20–50 copies of CMV DNA/mL of specimen. This chapter will focus on the performance of the currently available CMV molecular assays, specific quality control practices relevant to the clinical laboratory, and the clinical utility of molecular assays in managing patients with CMV disease. There have been many studies evaluating the clinical utility of molecular testing for diagnosis and monitoring of CMV disease. An exhaustive review of the literature is beyond the scope of this chapter; instead, we will focus on several key studies in transplant recipients and patients with AIDS.

2. AVAILABLE ASSAYS AND PERFORMANCE

Cytomegalovirus is a member of the Herpersviridae family and is a large double-stranded DNA enveloped virus, which replicates slowly and only grows in human diploid fibroblasts. Traditionally, the detection and quantitation of CMV has relied on culture and antigen-detection methods. Conventional culture methods, although considered the gold standard, are problematic because they are labor-intensive and have a slow turnaround time (1–3 wk). Rapid shell vial culture methods provide results in 1–2 d; however they lack sensitivity for detecting CMV in blood specimens *(1,2)*. The CMV antigenemia assay, which is widely used in clinical laboratories, is a semiquantitative nonculture method that detects the lower matrix protein, pp65, in polymorphonuclear cells. The assay has been shown to have a high sensitivity for the detection of CMV disease in solid organ transplant recipients *(1,3)* and results are available in 1–2 days. The major drawbacks of the assay are the lack of stability of the antigen in blood specimens and performing the test is labor-intensive. With these limitations of traditional assays, there has been great interest in developing molecular assays for use in CMV diagnostics.

The currently available molecular assays for the detection and quantification of CMV nucleic acid are listed in Table 1. At the time of this writing, the NucliSens assay and the qualitative version of the Hybrid Capture assay have been cleared by the Food and Drug Administration (FDA). As a group, these assays utilize a variety of amplification methods, including the polymerase chain reaction (PCR), nucleic acid sequence-based amplification (NASBA), and signal amplification (Hybrid Capture).

Polymerase chain reaction assays and the Hybrid Capture assay are designed to detect CMV DNA; they can be either qualitative or quantitative assays. On the other hand, the NASBA assay detects pp67 mRNA, which is transcribed late in the viral

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

Table 1
Molecular Assays for the Detection of CMV DNA

Assay	*Method*	*Target*	*Range*
AMPLICOR CMV MONITOR Test (Roce Diagnostics, Indianapolis, IN)	PCR[a]	CMV pol gene	400–100,000 copies/mL
Hybrid Capture System CMV DNA test (v.2) Digene Corporation, Gaithersburg, MD)	Hybrid Capture (signal amplification)	17% of CMV genome	1,400–6000 copies/mL)
NucliSens CMV Test (bioMerieux Inc., Durham, NC)	NASBA[b]	pp67 mRNA	Qualitative
In-house developed assays	Real-time PCR and conventional PCR	Variable	Variable

[a]PCR, polymerase chain reaction.
[b]NASBA, nucleic acid sequence-based amplification.

replicative cycle. Because the NASBA assay is isothermal, it can detect pp67 RNA without interference of CMV DNA. The NASBA assay is qualitative. The in-laboratory-developed PCR assays utilize both standard and real-time PCR methods and their performance varies as a result of differences in specimen type, nucleic acid extraction methods, target gene, primer sequence, quantitation standard, and detection method. These differences become quite important when comparing quantitative assays, as there might not be a close agreement between viral-load values obtained when testing the same sample in different assays. This makes it very difficult to compare results of studies using different assays and to establish viral-load values that correlate with clinical disease. The availability of commercial assays has been helpful in standardizing performance, although neither the quantitative Amplicor PCR test nor the quantitative Hybrid Capture assay have been cleared by the FDA.

The analytical performance of the Amplicor PCR assay has been evaluated in a study in which testing was done in three different laboratories using a plasma-based standard *(4)*. The standard was designed by making dilutions of electron-microscopy-counted viral particles into plasma. The linear range of the Amplicor PCR assay was established as 400 to 50,000 copies/mL of plasma. Recently, the manufacturer has modified the lower limit of the linear range to 600 copies/mL. The reproducibility of the assay varied depending on the copy number, with greater variability seen below 1000 copies/mL plasma. Overall, the standard deviation ranged from 0.11 to 0.48 $\log_{10}$. When the viral-load reached or exceeded 10,000 copies CMV DNA/mL, the standard deviation was consistently at or below 0.15 $\log_{10}$. Based on these data, it is likely that changes in plasma viral-load values will need to exceed 0.5–0.7 $\log_{10}$ copies/mL (threefold to fivefold) to represent clinically significant changes in viral replication. This is based on data from the Virology Quality Assessment Laboratory program (NIH/NIAID/DAIDS), in which a standard deviation of 0.15 $\log_{10}$ is used to ensure that a laboratory would maintain the precision required to have 90% power to detect a fivefold (0.69 $\log_{10}$) difference in copy number between two samples in the same batch *(5)*.

The analytical performance of the Amplicor PCR and the Hybrid Capture assays was recently compared using a cell-based standard *(6)*. This standard was developed by mixing known concentrations of CMV-infected fibroblasts with uninfected buffy coat cells. The Amplicor PCR assay could detect 100% of replicates when the CMV input concentration was 100 target cells. For the Hybrid Capture assay, 97% of replicates were detected at 3000 target cells, and the sensitivity was 100% at an input concentration of 10,000 target cells. Whether the 1.5 to 2.0 $\log_{10}$ increased analytical sensitivity seen for the Amplicor PCR assay compared to the Hybrid Capture is clinically significant requires further investigation. The specificity of the Amplicor assay was 100%, whereas the specificity of the Hybrid Capture was 93.8%. The linear range of the Amplicor PCR assay was 100 to 30,000 CMV target cells. The upper range of the Hybrid Capture assay was at least 100,000 target cells. Higher concentrations of target cells were not tested in the study. Overall, the variability of the Hybrid Capture assay was greater than that seen for the Amplicor assay. However, for both assays, the variability was greater near the lower limit of quantification. For concentrations of CMV target cells that were within the linear range of both assays, the viral-load values were within twofold (0.12 $\log_{10}$). These data are useful when comparing clinical results between the two assays using whole-blood specimens. However, many laboratories use the Amplicor PCR assay with plasma specimens and these data do not provide information on how viral-load values compare when testing plasma specimens with the Amplicor assay and whole-blood specimens with the Hybrid Capture assay. Viral-load values have been shown to be higher in polymorphonuclear cells than in plasma *(7–9)*. Similar results were observed in a study comparing CMV DNA levels obtained using plasma specimens in the Amplicor PCR assay and whole-blood specimens in the Hybrid Capture assay *(10)*. At the lower limit of detection of the Amplicor PCR assay (400 copies/mL), the corresponding Hybrid Capture results were 1.2 to1.3 $\log_{10}$ higher (7200 genomes/mL). On the other hand, when testing samples in the upper range of the Amplicor PCR assay, viral-load results between the two assays were similar, 100,000 and 170,000 copies/mL respectively for the Amplicor PCR and Hybrid Capture assays.

Another study has evaluated the viral-load levels in whole blood, plasma, peripheral blood leukocytes (PBLs), and peripheral blood mononuclear cells (PBMCs) using the Amplicor PCR assay *(9)*. A total of 319 samples were collected from 17 patients who developed 19 episodes of CMV disease. Viral-load values were significantly higher in the whole-blood compartment compared to plasma; the mean difference in plasma compared to whole-blood viral load was 0.67 $\log_{10}$ copies/mL. Using the same approach, there was no significant difference between PBLs and PBMCs. CMV DNA was detected more frequently in whole-blood specimens compared to plasma specimens. The

plasma-negative/whole-blood-positive specimens had a median viral load of 562 copies/mL. The authors stated that any of the sample types tested would be appropriate for testing during the time of CMV disease diagnosis. They preferred whole-blood specimens because of the higher sensitivity and higher viral-load levels obtained with this specimen, recognizing that low levels of CMV DNA in whole blood do not always correlate with disease. The study did not evaluate the utility of the different specimen types with regard to decisions regarding preemptive therapy. Both plasma and whole blood specimens are used for CMV testing in transplant recipients, and it is likely the viral-load level that correlates with clinical disease will differ based on the specimen tested. The results of the above study *(9)* indicate that the CMV DNA level correlating with active disease will likely be lower when using plasma specimens compared to whole-blood specimens.

There have been a few studies comparing the Amplicor PCR assay to in-laboratory-developed assays. One study *(11)* compared an in-laboratory-developed assay based on LightCycler technology with the Amplicor PCR assay in solid-organ transplant recipients with CMV infection. The Amplicor PCR assay yielded viral-load results that were 1 $\log_{10}$ higher than the LightCycler assay. In another study, a different in-laboratory-developed assay gave viral-load values 0.5 to 0.7 $\log_{10}$ higher than those seen for the Amplicor PCR assay *(4)*. These results underscore the variation seen in viral-load measurements between different in-laboratory-developed assays.

The Amplicor PCR and pp65 antigenemia assays compare very well *(12–14)* and provide valuable data for diagnosing active CMV disease, screening patients for use of preemptive therapy, and monitoring response to therapy. Fluctuations in CMV DNA levels in plasma and pp65 antigen in polymorphonuclear cells are very similar in both patients with active disease and those with asymptomatic infections. Although both assays provide very similar information, the viral-load values with the Amplicor PCR assay are 1 to 2 $\log_{10}$ higher than those observed with the antigenemia assay. This is not surprising considering the assays are measuring very different parameters. The antigenemia assay measures CMV pp65 antigen in polymorphonuclear cells, whereas the Amplicor PCR assay quantifies CMV DNA in plasma. The choice to use PCR testing or the antigenemia assay is often based on workflow, turnaround time, specimen stability, and cost.

Although many clinical laboratories use molecular assays to quantify CMV DNA, the interpretation of CMV viral-load testing remains a challenge because the cutoffs that correlate with clinical disease differ among assays and types of specimen. Also, it is likely that the viral load that correlates with disease might differ among transplant groups: bone marrow transplant recipients and different types of solid-organ transplant recipient. A key point to remember when reviewing the literature is to focus on studies that correlate CMV load with clinical outcome.

3. CLINICAL UTILITY

One of the primary uses of CMV molecular testing is for the diagnosis of active CMV disease. As mentioned, this involves distinguishing patients with clinically insignificant reactivation from those with symptomatic CMV disease. Molecular CMV assays are also commonly used to make decisions concerning the administration of preemptive antiviral therapy. With preemptive therapy, a group of patients at risk for developing CMV disease are further stratified based on the results of laboratory testing. For example, all members of the group would be tested for the presence of CMV DNA in their blood, and only those testing positive would be treated. Preemptive therapy focuses on a subgroup of patients determined to be at a higher risk of CMV disease. Therapy is administered prior to developing symptoms in an attempt to prevent the development of active disease. With prophylactic therapy, all patients in the group would be treated, without further stratification of risk. Thus, the use of prophylactic therapy involves treatment of a greater number of patients.

Once antiviral therapy is initiated, molecular CMV assays are useful for monitoring response to therapy. If there is inadequate response, antiviral resistance testing for CMV can be performed using a variety of molecular techniques.

One of the challenges in assessing the clinical performance of molecular CMV assays is developing an understanding that each assay measures a different parameter. The pp67 mRNA detected by the NucliSens assay is expressed at high level in patients with active CMV disease and is not expressed during latent infection *(15)*. Intracellular DNA present in both latently and actively infected cells is detected in whole blood or leukocytes by PCR assays and the Hybrid Capture assay. The CMV antigenemia assay measures pp65, a lower matrix antigen that is expressed late in the replicative cycle and is not detected in individuals with latent infection. Based on this information, it is likely that the NucliSens assay will be positive less frequently than CMV DNA PCR assays, using whole-blood or leukocyte specimens. In fact, this has been shown in several studies *(14,16)*. However, it is important to appreciate the difference between the analytical sensitivity and the clinical sensitivity of an assay. The analytical sensitivity of an assay refers to the number of viral particles or DNA molecules that can be detected, whereas clinical sensitivity is determined by comparing the assay to the diagnostic gold standard and requires an assessment of the clinical outcome of the patient. For example, an assay might have an excellent analytical sensitivity but not be useful in distinguishing patients with active CMV disease from those with latent infection or asymptomatic reactivation.

Early studies evaluating the utility of CMV DNA assays used very sensitive qualitative assays. As a result, CMV DNA was detected in patients without active disease *(17–19)*, thus limiting their clinical utility. Several different approaches have been used in an effort to improve the clinical specificity of molecular tests for CMV, including (1) detection of DNA in plasma rather than whole blood or leukocytes, (2) the use of assays that quantify CMV DNA, and (3) detection of mRNA rather than DNA. PCR assays that detect CMV DNA in plasma rather than leukocytes might provide a better correlation with clinical disease, as the detection of CMV DNA in plasma suggests active viral replication with spread of virus from leukocytes to the plasma factor *(20)*. Studies have shown that the detection of CMV DNA in plasma is an early marker for CMV disease in renal *(20)* and liver *(21)* transplant recipients. The goal with the use of quantitative DNA assays is to establish threshold viral-load levels that can distinguish latent infection and asymptomatic reactivation from active CMV disease. As

discussed earlier, the advantage of measuring pp67 mRNA is that it is not expressed in latently infected cells.

3.1. TRANSPLANT RECIPIENTS Studies of transplant recipients using in-laboratory-developed quantitative PCR assays have shown that patients with active CMV disease have higher viral-load values than those with asymptomatic infection *(22–25)*. However, because of differences in assay design, the cutoff value associated with active disease was different for each assay. Although these assays have proven to be very useful for individual laboratories, it has been difficult to establish broadly applicable, clinically relevant guidelines for the use of in-laboratory-developed viral load assays. Recently, several studies have evaluated the clinical utility of two standardized commercially available assays: the Amplicor PCR test and the Hybrid Capture assay. Humar et al. *(12)* enrolled 97 liver transplant recipients and followed them for 12 wk posttransplant with weekly plasma viral-load testing (Amplicor) and antigenemia testing. Of the 97 patients enrolled, 61 had CMV infection, which was defined as a positive laboratory test. Twenty-one patients developed CMV disease, including hepatitis, CMV viral syndrome, and esophagitis or colitis. As shown with early studies, patients with CMV disease had a higher peak viral load than those with asymptomatic infection. The median peak viral load in patients with asymptomatic CMV infection was 1820 copies/mL compared to 55,000 copies/mL for those with symptomatic CMV disease. The value of viral-load testing in predicting CMV disease was also assessed. Using receiver-operating characteristic curves, the authors determined the optimal viral-load for detecting CMV disease. As the viral-load cutoff level for detecting CMV disease was increased, there was a loss in sensitivity, but an improvement in specificity was observed. For example, using a cutoff viral load of 400 copies/mL, the sensitivity for detecting CMV disease was 100%, but the specificity was only 47%, with a positive and negative predicative value of 34% and 100%, respectively. With a viral-load cutoff of ≥5000 copies/mL, the sensitivity decreased to 86%, but the specificity increased to 87%, with a positive and negative predictive value of 64% and 96%, respectively. Based on the data it was determined that a viral load of 2000–5000 copies/mL was the most predictive of CMV disease. A similar study was conducted in renal transplant recipients using the Hybrid Capture assay to quantify CMV DNA from whole-blood specimens *(17)*. The risk of CMV disease increased from 1.5% when the viral load was 10,000 copies/mL blood up to 73% when the viral load was 1,000,000 copies/mL blood. Both of these studies provide very useful information regarding when to initiate anti-CMV preemptive therapy in solid organ transplant recipients. The viral-load cutoffs for these assays differ because the Amplicor PCR assay is measuring CMV DNA load in plasma and the Hybrid Capture assay is quantifying CMV DNA in whole blood. Also, viral-load cutoffs established for one type of solid organ transplant might not apply to all types of solid organ transplant.

There are several ongoing randomized controlled trials evaluating the utility of the NucliSens assay for decisions regarding initiation of preemptive therapy. These studies are comparing the NucliSens assay to either antigenemia or PCR in solid organ or bone marrow transplant recipients. The results of these clinical trials will provide key information regarding the clinical utility of the NucliSens assay. An earlier study has shown the NucliSens assay to be positive less frequently than the Amplicor PCR test, but it was positive in 12 of 13 episodes of active CMV disease *(14)*. In the one episode of disease in which the NucliSens assay was negative, the patient was on ganciclovir therapy prior to collecting the specimen. Many of the samples that were Amplicor PCR positive and NucliSens negative occurred in patients on therapy or without symptoms. The positivity rate of the NucliSens assay was related to CMV DNA levels. Ninety-seven percent of the specimens with a viral load > 5000 copies/mL in the Amplicor PCR assay were positive in the NucliSens assay. When the viral load was ≤5000 copies/mL, only 32% of the specimens were positive in the NucliSens assay. This data further support the concept that it is important to assess the NucliSens assay based on clinical disease, not just in comparison to a sensitive PCR assay.

Molecular assays are also useful in monitoring response to antiviral therapy in transplant recipients. Aitken et al. reported a greater than 90% reduction in viral load, measured with the Hybrid Capture assay, after initiation of therapy for CMV infection. Several studies using the Amplicor PCR assay have reported that viral loads become undetectable several weeks after initiating therapy *(12–14)*. In fact, failure of viral load to decline after a few weeks of appropriate therapy is a cause for concern, as patients with documented ganciclovir resistance have been shown to have persistently elevated viral load levels while on therapy *(13)*. In these patients, the viral load was 20,000 to 70,000 copies/mL as measured with the Amplicor PCR assay.

Another application of molecular assays is identifying patients at risk for relapsing CMV infection. In one study *(26)*, 24 solid organ transplant recipients with CMV infection had viral loads measured by the Amplicor PCR test prior to and after the completion of 14 d of ganciclovir therapy. Eight of the 24 patients experienced relapsing CMV infection. Although the pretreatment viral load did not predict those at risk for relapse, the presence of a detectable viral load after completion of ganciclovir therapy was a risk factor for relapsing infection. Patients in the nonrelapse group had a median posttreatment plasma viral load of <400 copies/mL, whereas those in the relapse group had a median post-treatment viral load of 29,200 copies/mL ($p = 0.007$). Similar results have been reported with the NucliSens assay *(15);* all 6 solid organ recipients with detectable RNA after completing a course of therapy required further therapy for their CMV infection.

The rate of decline in CMV load has also been shown to be predictive of relapsing CMV infections. Fifty-two organ transplant recipients with CMV disease were monitored with the Amplicor PCR test, and 12 developed recurring CMV disease. The time to clearance of CMV DNA from plasma was 33.8 d in the group with recurrent CMV disease compared to 17.2 d in the group without recurrent disease ($p = 0.002$) *(27)*. The viral-load half-life was 8.8 d compared to 3.2 d ($p = 0.001$) in the group with and without recurrent disease, respectively. Following viral loads closely after initiating therapy will allow the identification of patients at risk of recurrent infection, thus providing the opportunity to intensify therapy and possibly prevent recurrent infection.

In summary, molecular assays that detect mRNA or quantify CMV DNA have been shown to be useful in diagnosing active CMV disease, making decisions regarding preemptive therapy, and monitoring response to therapy. The viral-load level that correlates with active disease, the risk of active disease,or the risk of relapse will likely vary for the different assays used.

3.2. HIV-1 INFECTION There is increasing evidence that the risk of developing CMV disease in persons with AIDS is directly related to the quantity of CMV DNA in plasma *(28,29)*. An early study by Spector et al. *(28)* evaluated the utility of CMV DNA load in predicting CMV disease and survival in AIDS patients. The specimens for the study were collected from human immunodeficiency virus (HIV)-1-infected persons participating in a clinical trial to determine the efficacy of oral ganciclovir for preventing CMV disease. Among placebo recipients, detection of CMV DNA in plasma was associated with a 3.4-fold increased risk of developing CMV disease and a 2.5-fold increased risk of death. Each $\log_{10}$ increase in baseline CMV DNA load was associated with a 3.1-fold increase in CMV disease. A subsequent study *(29)* showed that in patients with advanced AIDS, CMV DNA load is an independent marker of CMV disease and survival and CMV DNA load is more predictive of these outcomes than HIV-1 viral load. In addition, patients with detectable CMV DNA in plasma who respond to preemptive therapy with oral ganciclovir, defined as an undetectable CMV DNA in plasma, had a significantly lower risk of developing CMV disease and higher rates of survival, despite stable or increasing HIV-1 viral-load levels. Similar results were shown in a study *(30)* of patients with advanced AIDS, where the detection of CMV DNA in plasma was a significant risk factor for CMV disease progression. Despite these findings, AIDS patients were not routinely monitored with CMV viral-load assays to assess the risk of developing end-organ CMV disease. This was the result of several factors: (1) concern that use of oral ganciclovir as preemptive therapy might lead to the development of ganciclovir-resistant CMV; (2) the sharp decline in the incidence of CMV end organ disease with the use of highly active antiretroviral therapy (HAART) *(31,32)*.

A more recent study *(33)* of CMV end-organ disease in 403 persons with advanced HIV-1 infection evaluated the utility of screening patients for CMV DNA in blood fractions. The vast majority of the participants in the study (84%) were receiving HAART, which more closely resembles current clinical practice. Of the 21 individuals who developed CMV end-organ disease, 20 had ≤50 CD_4 cells/mm^3 and an HIV-1 viral load of >10,000 copies/mL plasma. Among the 107 subjects with ≤50 CD_4 cells/mm^3 at baseline, none of the 19 who had an HIV-1 viral load of ≤10,000 copies/mL at baseline developed CMV end-organ disease. In those individuals with ≤50 CD_4 cells/mm^3 and >10,000 HIV-1 RNA copies/mL plasma at baseline, a rise in CMV DNA above the detection limit was associated with the development of CMV end-organ disease. CMV viral-load testing for this study was done with both the Amplicor PCR assay (plasma specimens) and the Hybrid Capture assay (whole-blood specimens). There were more specimens positive with the Hybrid Capture assay compared to the Roche assay; however, both assays were useful in establishing the risk of CMV disease. The results of this study provided evidence that screening for CMV DNA could be focused on AIDS patients with ≤50 CD_4 cells/mm^3 and an HIV-1 viral load of >10,000 copies/mL plasma. If CMV DNA was detected, the use of preemptive therapy to prevent CMV end-organ disease could be considered. This approach would allow the targeted use of both CMV molecular testing, as well as anti-CMV therapy.

Neurological disease from CMV occurs in immunocompromised individuals. Common manifestations include encephalitis/ventriculo-encephalitis, retinitis, and polyradiculopathy. Culture of CMV from the cerebrospinal fluid (CSF) is very specific but the sensitivity is no better than 50–60% for the diagnosis of neurologic disease resulting from CMV *(34)*. For this reason, CMV DNA PCR assays have been investigated as a diagnostic tool for identifying neurologic CMV disease. Most studies have utilized qualitative PCR assays. Several groups have summarized the literature and reported that CMV DNA PCR assays are quite sensitive (80–100%) for the diagnosis of CMV neurologic disease, whereas the specificity varied from 75% to 100% *(34,35)*. One study *(36)* reported the sensitivity of CMV DNA PCR to be 91%; however, the specificity was only 42% when compared to neuropathologic examination. In this study, the CSF samples analyzed were obtained at the time of autopsy, which might explain the poor specificity. Most studies of CSF specimens collected prior to death have shown CMV DNA PCR to have a much higher specificity. Thus, CMV DNA PCR assays using CSF specimens have proven to be useful for the diagnosis of CMV-associated neurologic disease, even though CMV DNA might occasionally be detected in the CSF of individuals without disease.

The clinical correlation between the detection of CMV pp67 mRNA in CSF and active CMV central nervous system (CNS) disease was evaluated in patients with HIV-1 infection *(37)*. A total of 76 CSF specimens from 65 individuals with and without CMV CNS disease were tested by the NucliSens assay and a nested CMV DNA PCR. Viral culture results were available for 54 of the 76 specimens. Resolution of discordant results was based on clinical, laboratory, and radiographic studies and a review of antiviral therapy. The specificity of the nested PCR, NucliSens pp67 assay, and culture were 90.5%, 100%, and 100%, respectively, for the diagnosis of CMV CNS disease. Of the three assays, the nested DNA PCR had the highest sensitivity (100%), followed by the NucliSens (84.6%), and then culture (18%). The detection of pp67 mRNA in CSF had a good correlation with active CMV CNS disease, whereas CSF culture was insensitive. Consistent with the previous studies, the CMV DNA PCR was very sensitive, but occasionally detected CMV DNA in patients without CMV CNS disease.

Because of the insensitivity and slow turnaround of culture, many laboratories now rely on molecular assays for the diagnosis of CMV-associated neurologic disease. This appears to be a reasonable approach, based on overall assay performance, although caution must be used when interpreting results resulting from the limitations in sensitivity for the NucliSens assay and specificity for DNA PCR assays.

4. CLINICAL LABORATORY ISSUES

In general, a variety of clinical specimens are used in CMV molecular assays, including plasma, leukocytes, whole blood, and CSF. However, individual assays could have specific spec-

imen requirements. For example, the Amplicor PCR assay has been developed for plasma specimens (200 µL), but the assay can be adapted for leukocytes, whole blood, or CSF specimens. The Hybrid Capture assay uses 1.5 mL of whole blood, whereas the NucliSens has been designed for 100 µl of whole blood, although larger volumes of blood can be used. CSF specimens can also be used in the NucliSens assay *(37)*. For both the Amplicor PCR and Hybrid Capture assays, EDTA is the preferred anticoagulant, whereas the NucliSens assay can accommodate any anticoagulant.

An advantage of PCR assays is that CMV DNA is stable in whole blood or plasma for 3–4 d when stored at room temperature *(38,39)* and for up to 5 d when stored at 4°C *(13,39)*. Another advantage of CMV DNA assays that utilize plasma specimens is that they can be performed on patients with low leukocyte counts. On the other hand, the antigenemia assay requires 3–5 mL of blood, which should be processed within 8 h of collection to avoid a deterioration of the antigen, and the assay is adversely affected by low leukocyte counts.

Maintaining adequate quality control of CMV molecular assays can be challenging because of the lack of FDA-approved quantitative assays and the limited availability of control material and proficiency testing. External CMV control material is now available from Advanced Biotechnologies, Inc. (Columbia, MD), Boston Biomedica, Inc. (West Bridgewater, MA), and Aero Metrix (Benicia, CA) Currently, proficiency testing for qualitative CMV assays is available from the College of American Pathology Surveys program. In addition, a Quality Control for Molecular Diagnostics program organized in collaboration with the European Society for Clinical Virology and the European Society for Clinical Microbiology and Infectious Diseases provides proficiency testing for quantitative CMV assays.

5. RESISTANCE TESTING

There are currently three antiviral drugs that are FDA approved for the treatment of systemic CMV disease. These are ganciclovir (GCV), cidofovir (CDV) and foscarnet (phosphonoformic acid [FOS]). One additional drug, fomivirsen, has been approved for intravitreal treatment of CMV retinitis *(40);* however, its use has been limited. GCV is the most widely used of all of these drugs. However, GCV itself is not active but requires an initial phosphorylation step by a kinase (phosphotransferase) encoded by the CMV UL97 gene (Fig. 1) *(41,42)*. Cellular enzymes add two more phosphates to make the active form, which is an analog of guanosine triphosphate. This analog is incorporated into the replicating DNA genome by the CMV DNA polymerase (encoded by the UL54 gene), which results in termination of replication *(41,42)*. Thus, there are two virally encoded enzymes that interact with GCV: the UL97 kinase and the UL54 DNA polymerase.

Cidofovir is a cytidine monophosphate analog, which does not require a viral enzyme for activation (Fig. 1). Cellular enzymes add additional phosphates to produce the triphosphate analog, which is then incorporated into the replicating DNA by the CMV polymerase. The result is termination of DNA replication, but, in this case, the polymerase is the only viral product required for antiviral activity *(41,42)*. The third drug used for systemic CMV infections is FOS, which is a pyrophosphate analog. It does not require any viral or cellular enzymes for activation (Fig. 1), and it targets the pyrophosphate binding site on the CMV DNA polymerase *(41,42)*. Again, DNA replication is inhibited, but the mechanism differs from that of GCV and CDV.

The majority of patients requiring antiviral therapy for CMV infection are transplant recipients or HIV-1 infected patients *(42)*. Both groups are immunocompromised, which means that lengthy antiviral treatment is required when CMV disease arises. Long-term treatment regimens are associated with antiviral drug resistance. This is because CMV is a DNA virus replicated by a DNA polymerase, and DNA polymerases have a proof-reading function that prevents high mutation rates. Thus, selection of low numbers of drug-resistant virus from the initial wild-type drug-sensitive population generally only occurs after weeks to months of therapy. There are two different types of assay for the detection of drug-resistant virus isolates: phenotypic and genotypic.

Phenotypic assays rely on analysis of growth characteristics of virus isolates in the presence of different concentrations of antiviral drugs and, therefore, require viable virus. The gold standard among phenotypic assays is the plaque reduction assay (PRA) *(43)*. Permissive cell monolayers (most often human foreskin fibroblasts) are infected with enough virus to produce countable discrete foci of infection or plaques in the absence of drug. Different concentrations of the drug of interest are added to the overlay medium of sets of replicate wells. After 10–14 d of incubation, plaques are counted and averaged for each concentration. The concentration that reduces the number of plaques by 50% relative to control wells without drug is defined as the 50% inhibitory concentration (IC_{50}). Experience with assays for each of the antiviral drugs has led to the use of cutoffs to determine susceptibility to each drug. Generally accepted IC_{50} values for defining resistance by the PRA are GCV > 6 µ*M*, CDV > 2 µ*M*, and FOS > 400 µ*M* *(44,45)*.

The main problems in performing the PRA are that it is time-consuming, labor-intensive, and subjective. It generally takes 1–3 wk for the clinical laboratory to identify CMV in culture, 1–2 wk to grow a sufficient amount of virus to inoculate the assay, and an additional 1–2 wk before the assay can be read. Counting plaques is problematic, because it is difficult to distinguish truly resistant plaques from plaques containing drug-sensitive virus that are resolving in the presence of drug.

Other phenotypic assays have been developed to address some of these problems. A commercially available DNA hybridization assay, Hybriwix (Diagnostic Hybrids, Athens, OH), is based on quantitation of radiolabeled probe binding to viral DNA following replication in cell culture in the presence of different concentrations of drug *(46,47)*. The results are expressed as an IC_{50}, but, in this case, the IC_{50} represents a 50% reduction in radioactivity counts in the presence of drug compared to the control without drug. This method only slightly shortens the assay, but it reduces the labor and eliminates the subjectivity of the PRA. The use of radioactivity is a drawback. A flow-cytometry-based assay is based on detection of fluorescent antibody binding to an immediate early CMV protein in monolayers exposed to different drug concentrations *(48)*. IC_{50} results are determined by comparison of numbers of fluorescent cells produced in the presence of each

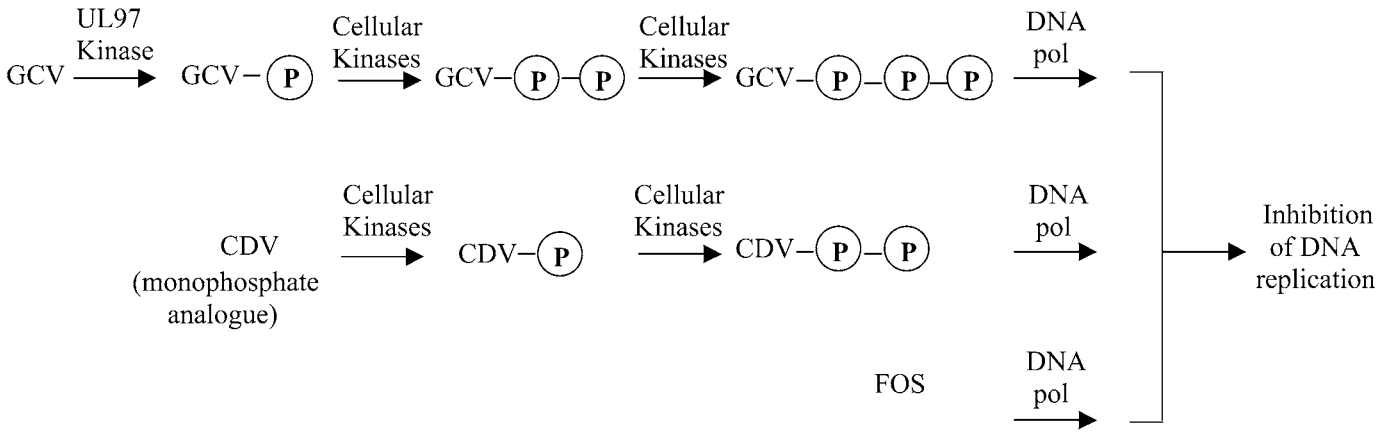


Fig. 1. Antiviral activity of anti-CMV drugs. Phosphate groups are designated by a circled P. GCV, ganciclovir; CDV, cidofovir; FOS, foscarnet or phosphonoformic acid; DNA pol, DNA polymerase. (See text for explanation.).

drug concentration with the number produced by controls without drug as read by a flow cytometer. Again, the assay is somewhat shortened compared to the PRA and results are not subjective.

Despite such efforts to improve the phenotypic assay, they all require at least a month to obtain results, which is generally not therapeutically useful. The lengthy time requirement has led to development of genotypic assays. These assays are based on detection of nucleotide sequence changes that are found in drug-resistant CMV strains. The currently used antiviral drugs target the UL97 gene (GCV) and/or the DNA polymerase gene (GCV, CDV, FOS). Evidence has accumulated to show that point mutations or short deletions in these genes confer drug resistance.

Because GCV is the most widely used anti-CMV drug, GCV resistance mutations are most frequently mapped to the UL97 gene. These mutations are well-defined at three specific sites, within a 700-nucleotide region at the 3′ end of the gene. These are point mutations within codons 460 and 520 and either point mutations or deletions within the codon range 590–607 *(47,49,50)*.

Mutations in the DNA polymerase gene appear to occur less frequently than those in the UL97 gene *(41,42,51)*. The majority of polymerase mutations conferring resistance to GCV are also cross-resistant to CDV *(41)*. In general, FOS resistance mutations are not cross-resistant to either of the other two drugs, although a double deletion of codons 981 and 982 has been reported to confer resistance to all three drugs *(52)*. There are at least 23 sites within this gene where drug resistance mutations have been mapped *(51)*. This covers approx 2000 nucleotides, a much larger region than that of the UL97 gene.

Initial genotypic assays to detect some of the more common UL97 mutations were based on loss or gain of restriction enzyme cleavage sites *(42,47)*. The difficulty with this approach is that not all of the presently confirmed resistance mutations alter known restriction enzyme recognition sites. In addition, base changes not associated with drug resistance can produce new restriction sites near the known resistance sites and lead to misinterpretation of results.

A more recent approach for detection of resistance mutations is genotypic analysis based on automated sequencing, which gives much more definitive results than restriction enzyme analysis. CMV DNA can be extracted from cultures from the clinical laboratory and used as a template for PCR amplification of products containing the nucleotide regions known to contain mutations *(41,45)*. Sequencing of the PCR products can identify drug-resistant mutant strains as well as mixtures of mutant and wild-type drug-sensitive strains. Baseline sequences of both the UL97 and DNA polymerase genes from drug-sensitive strains have been recently published. These sequences serve to identify the natural sequence variability of the two genes that is unrelated to drug resistance *(53,54)*.

Further improvement of the sequencing assay has led to the development of methods in which CMV DNA is extracted directly from patient specimens. The target regions of the DNA are amplified by PCR and the PCR products are sequenced to detect drug-resistance mutations. Direct sequencing of patient specimens eliminates lengthy culture. The time required to identify drug-resistant virus is then reduced to less than a week from specimen collection. This rapid method can, therefore, provide definitive results within a therapeutically useful time frame.

It should be noted that for a mutation to be classified as a drug-resistance mutation, drug-resistance has to be confirmed by transfer of that mutation to a drug-sensitive strain *(51)*. The list of marker transfer-confirmed mutations for both the UL97 and DNA polymerase (UL54) genes is now extensive (Table 2). One caveat for genotypic assays is the fact that there is some normal baseline sequence variability in the UL97 and DNA polymerase genes in drug-sensitive strains *(53,54)*. The assumption is made that the same mutation will produce resistance in all genetic backgrounds, which might not prove to be the case in the future as more strains are examined. This brings us back to the fact that phenotypic assays are biologically more relevant than genotypic assays for detection of drug resistance, because individual strains are tested for replication directly in the presence of antiviral drug. However, the experience so far suggests that genotypic results correlate with the clinical prediction of drug resistance and provide rapid results not possible for the presently available phenotypic assays *(41,42,53)*.

6. FUTURE DIRECTIONS

Over the past several years, a wide variety of studies have established the clinical utility of molecular CMV assays. As a result, these tests now play a central role in managing patients at risk of developing CMV reactivation and those with active

Table 2
CMV Drug Resistance Mutations

Codon	*Amino wild-type*	*Amino mutant*	*GCV*	*CDVR*	*FOSR*
UL97 mutations					
460	Met	Val/Ile	+	–	–
520	His	Gln	+	–	–
591	Ala	Val	+	–	–
592	Cys	Gly	+	–	–
594	Ala	Thr/Val	+	–	–
591–594		Deletion	+	–	–
595–603		Deletion	+	–	–
595	Leu	Phe/Ser/Trp	+	–	–
595		Deletion	+	–	–
596	Glu	Gly	+	–	–
598	Gly	Ser	+	–	–
599	Lys	Thr	+	-	-
600		Deletion	+	–	–
601	Thr	Met	+	–	–
603	Cys	Trp	+	–	–
607	Cys	Phe/Tyr	+	–	–
DNA Polymerase Mutations					
301	Asp	Asn	+	+	–
408	Asn	Asp	+	+	–
410	Asn	Lys	+	+	–
412	Phe	Cys	+	+	–
413	Asp	Glu	+	+	–
501	Leu	Ile	+	+	–
503	Thr	Ile	+	+	–
513	Lys	Glu/Asn	+	+	–
516	Leu	Arg	+	+	–
522	Pro	Ser	+	+	–
545	Leu	Ser	+	+	–
588	Asn	Asp	–	–	+
700	Thr	Ala	–	–	+
715	Val	Met	–	–	+
756	Glu	Lys/Asp/Gln	–	–	+
781	Val	Ile	–	–	+
787	Val	Ile	–	–	+
802	Leu	Met	+	–	+
809	Ala	Val	+	–	+
812	Val	Leu	+	–	+
821	Thr	Ile	+	–	+
981–982		Deletion	+	+	+
987	Ala	Gly	+	+	–

*Superscript R designates resistance.

disease. Looking to the future, real-time molecular assays, which allow the amplification and detection steps to occur simultaneously, will expand the clinical utility of molecular CMV assays. These real-time assays decrease the time required to perform the assays and reduce the risk of contamination because of of carryover of amplicons because there is no post-amplification processing. They are very sensitive assays, which provide a 5–6 $\log_{10}$ linear range. These advances will further increase the value of molecular CMV assays for both the clinical laboratory and the clinicians. With the availability of rapid methods, resistance testing will also play an increasingly important role in clinical management decisions. The standardization of automated sequencing methods and the characterization of mutations associated with GCV, CDV, and FOS resistance will make it feasible for clinical laboratories to routinely offer genotypic resistance testing. A key advance of genotypic resistance testing is the availability of results in a time frame that impacts clinical care.

There are several major challenges that remain for laboratories offering molecular CMV assays. Well-characterized CMV DNA standards are needed so there is agreement of viral-load values obtained with commercial and in-laboratory-developed assays. Once this occurs, viral-load cutoffs that correlate with clinical disease can be established that will be independent of the assay used to generate the result. There is also a need for a more comprehensive proficiency program for CMV load assays and resistance testing, which will likely occur as molecular CMV assays are used in a greater number of clinical laboratories.

REFERENCES

1. Mazzulli, T., Rubin, R. H., Ferraro, M. J., et al. Cytomegalovirus antigenemia: clinical correlations in transplant recipients and in persons with AIDS. *J. Clin. Microbiol.* **31:**2824–2827, 1993.
2. Mazzulli, T., Drew, L., Yen-Lieberman, B., et al. Multicenter comparison of the Digene Hybrid Capture CMV DNA assay (version 2.0), the pp65 antigenemia assay, and cell culture for detection of cytomegalovirus viremia. *J. Clin. Microbiol.* **37(4):**958–963, 1999.
3. van den Berg, A. P., Klompmaker, I. J., Haagsma, E. B., et al. Antigenemia in the diagnosis and monitoring of active cytomegalovirus infection after liver transplantation. *J. Infecti. Dis.* **164:**265–270, 1991.
4. Caliendo, A. M., Schuurman, R., Yen-Lieberman, B., et al. and The CMV Working Group of the Complications of HIV Disease RAC AIDS Clinical Trials Group, Comparison of quantitative and qualitative PCR assays for cytomegalovirus DNA in plasma. *J. Clin. Microbiol.* **39(4):**1334–1338, 2001.
5. Yen-Lieberman, B., Brambilla, D., Jackson, B., et al. Evaluation of a quality assurance program for quantification of human immunodeficiency virus type 1 RNA in plasma and AIDS Clinical Trials Group virology laboratories. *J. Clin. Microbiol.* **34:**2695–2701, 1996.
6. Caliendo, A. M., Yen-Lieberman, B., Baptista, J., et al. Comparison of molecular tests for detection and quantification of cell-associated CMV DNA. *J. Clin Microbiol.* **41:**3509–3513, 2003.
7. Boivin, G., Handfield, J., Toma, E., Murray, G., Lalonde, R., and Bergeron, M. G. Comparative evaluation of the cytomegalovirus DNA load in polymorphonuclear leukocytes and plasma of human immunodeficiency virus-infected subjects. *J. Infecti. Dis.* **177:**355–360, 1998.
8. Gerna, G., Zavattoni, M., Baldanti, F., et al. Human cytomegalovirus (HCMV) leukoDNAemia correlates more closely with clinical symptoms than antigenemia and viremia in heart and heart-lung transplant recipients with primary HCMV infection. *Transplantation* **65(10):**1378–1385, 1998.
9. Razonable, R. R., Brown, R. A., Wilson, J. A., et al. The clinical use of various blood compartments for cytomegalovirus (CMV) DNA quantitation in transplant recipients with CMV disease. *Transplantation* **73(6):**968–973, 2002.
10. Tong, C. Y. W., Cuevas, L. E., Williams, H., and Bakran, A. Comparison of two commercial methods for measurement of cytomegalovirus load in blood samples after renal transplantation. *J. Clini. Microbiol.* **38:**1209–1213, 2000.
11. Razonable, R. R., Brown, R. A., Espy, M. J., et al. Comparative quantitation of cytomegalovirus (CMV) DNA in solid organ transplant recipients with CMV infection by using two high-throughput automated systems. *J. Clini. Microbiol.* **39:**4472–4476, 2001.
12. Humar, A., Gregson, D., Caliendo, A. M., et al. Clinical utility of quantitative cytomegalovirus viral load determination for predicting cytomegalovirus disease in liver transplant recipients. *Transplantation* **68:**1305–1311, 1999.
13. Caliendo, A. M., St. George, K., Kao, S.-Y., et al. Quantitative CMV PCR in plasma as compared with CMV antigenemia: clinical utility of the prototype AMPLICOR CMV MONITOR test in transplant recipients. *J. Clini. Microbiol.* **38:**2122–2127, 2000.

14. Caliendo, A. M., St. George, K., Allega, J., Bullotta, A. C., Gilbane, L., and Rinaldo, C. R. Distinguishing cytomegalovirus (CMV) infection and disease with CMV nucleic acid assays. J. Clin. Microbiol. **40:**1581–1586, 2002.
15. Oldenburg, N., Lam, K., Khan, M., et al. Evaulation of human cytomegalo-virus gene expression in thoracic organ transplant recipients using nucleic acid sequence-based amplification. *Transplantation* **70(8):** 1209–1215, 2000.
16. Gerna, G., Baldanti, F., Middeldorf, J., et al. Clinical significance of expression of human cytomegalovirus pp67 late transcript in heart, lung, and bone marrow transplant recipients as determined by nucleic acid sequence-based amplification. *J. Clin. Microbiol.* **37(4):**902–911, 1999.
17. Aitken, C., Barrett-Muir, W., Millar, C. T., et al. Use of molecular assays in diagnosis and monitoring of cytomegalovirus disease following renal transplantation. *J. Clini. Microbiol.* **37:**2804–2807, 1999.
18. Gerna, G., Zipeto, D., Parea, M., et al. Monitoring of human cytomegalovirus infections and ganciclovir treatment in heart transplant recipients by determination of viremia, antigenemia and DNAemia. *J. Infecti. Dis.* **164:**488–498, 1991.
19. Tanabe, K., Tokumoto, T., Ishikawa, N. et al. Comparative study of cytomegalovirus (CMV) antigenemia assay, polymerase chain reaction, serology, and shell vial assay in the early diagnosis and monitoring of CMV infection after renal transplantation. *Transplantation* **64:**1721–1725, 1997.
20. Wolf, D. and Spector, S. Early diagnosis of human CMV disease in transplant recipients by DNA amplification. *Transplantation* **56:**330–334, 1993.
21. Patel, R., Smith, T. F., Espy, M., et al. A prospective comparison of molecular diagnostic techniques for the early detection of cytomegalovirus in liver transplant recipients. *J. Infecti. Dis.* **176:**1484–1490, 1995.
22. Weinberg, A., Hodges, T. N., Li, S., Cai, G., and Zamora, M. R. Comparison of PCR, antigenemia assay, and rapid blood culture for detection and prevention of cytomegalovirus disease after lung transplantation. *J. Clini. Microbiol.* **38:**768–772, 2000.
23. Roberts, T. C., Brennan, D. C., Buller, R. S., et al. Quantitative polymerase chain reaction to predict occurrence of symptomatic cytomegalovirus infection and assess response to ganciclovir therapy in renal transplant recipients. *J. Infecti. Dis.* **178:**626–635, 1998.
24. Ferreira-Gonzalez, A., Fisher, R. A., Weymouth, L. A., et al. *C*linical utility of a quantitative polymerase chain reaction for diagnosis of cytomegalovirus disease in solid organ transplant patients. *Transplantation* **68:**991–996, 1999.
25. Cope, A. V., Sabin, C., Burroughs, A., Rolles, K., Griffiths, P. D., and Emery, V. C. Interrelationships among quantity of human cytomegalovirus (HCMV) DNA in blood, donor–recipient serostatus, and administration of methylprednisolone as risk factors for HCMV disease following liver transplantation. *J. Infecti. Dis.* **176:**1484–1490, 1997.
26. Sia, I. G., Wilson, J. A., Groettum, C. M., Espy, M. J., Smith, T. F., and Paya, C. V. Cytomegalovirus (CMV) DNA load predicts relapsing CMV infection after solid organ transplantation. *J. Infecti. Dis.* **181:**717–720, 2000.
27. Humar, A., Kumar, D., Boivin, G., and Caliendo, A. M. Cytomegalovirus (CMV) viral load kinetics to predict recurrent disease in solid organ transplant patients with CMV disease. *J. Infecti. Dis.* **186:**829–833, 2002.
28. Spector, S. A., Wong, R., Hsia, K., Pilcher, M., and Stempien, M. J. Plasma cytomegalovirus (CMV) DNA load predicts CMV disease and survival in AIDS patients. *J. Clini. Investi.* **101(2):**497–502, 1998.
29. Spector, S. A., Hsia, K., Crager, M., Pilcher, M., Cabral, S., and Stempien, M. J. Cytomegalovirus (CMV) DNA load is an independent predictor of CMV disease and survival in advanced AIDS. *J. Virol.* **73(8):**7027–7030, 1999.
30. Nokta, M., Holland, F., De Gruttola, V., et al. Cytomegalovirus (CMV) polyerase chain reaction profiles in individuals with advanced human immunodeficiency virus infection: relationship to CMV disease (Protocol 204/Glaxo Wellcome 123-014, International CMV Prophylaxis Study Group). *J. Infecti. Dis.* **185:**1717–1722, 2002.
31. Palella, F. J. J., Delaney, K. M., Moorman, A. C., et al. and H.O.S. Investigators, Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. *N. Engl. J. Med.* **338(13):**853–860, 1998.
32. Casado, J. L., Arrizabalaga, J., Montes, M. et al. Incidence and risk factors for developing cytomegalovirus retinitis in HIV-infected patients receiving protease inhibitor therapy. *AIDS* **13(12):** 1497–1502, 1999.
33. Erice, A., Tierney, C., Hirsch, M., et al. CMV and HIV viral burden and development of CMV end-organ disease: a prospective study in HIV-infected subjects (ACTG 360) in 9th Annual Conference on Retroviruses and Opportunistic Infections. Seattle, WA, 2002. (abstract).
34. Arribas, J. R., Storch, G. A., Clifford, D. B., and Tselis, A. C. *Cytomegalovirus encephalitis. Ann. Intern. Med.* **125(7):**577–587, 1996.
35. McCutchan, J. A. Clinical impact of cytomegalovirus infections of the nervous system in patients with AIDS. *Clin. Infecti. Dis.* **21(Suppl. 2):**S196–S201, 1995.
36. Achim, C. L., Nagra, R. M., Wang, R., Nelson, J. A., and Wiley, C. A. Detection of cytomegalovirus in cerebrospinal fluid autopsy specimens from AIDS patients. *J. Infecti. Dis.* **169(3):**623–627, 1994.
37. Zhang, F., Tetali, S., Wang, X. P., Kaplan, M. H., Cromme, F., and Ginocchio, C. Detection of human cytomegalovirus pp67 late gene transcripts in cerebrospinal fluid of human immunodeficiency virus type 1-infected patients by nucleic acid sequence-based amplicafication. *J. Clin. Microbiol.* **38(5):**1920–1925, 2000.
38. Schafer, P., Tenschert, W., Gutensohn, K., and Laufs, R. Minimal effect of delayed sample processing on results of quantitative PCR for cytomegalovirus DNA in leukocytes compared to results of an antigenemia assay. *J. Clin. Microbiol.* **35:**741–744, 1997.
39. Yen-Lieberman, B., Starkey, C., and Kohn, D. The effect of storage conditions on the stability of CMV DNA in whole blood and plasma specimens. in American Society of Microbiology 101st Meeting, Orlando, FL, 2001 (abstract).
40. Naesens, L. and De Clercq, E. Recent developments in herpesvirus therapy. *Herpes* **8(1):**12–16., 2001.
41. Chou, S. Antiviral drug resistance in human cytomegalovirus. *Transplant. Infecti. Dis.* **1:**105–114, 1999.
42. Erice, A. Resistance of human cytomegalovirus to antiviral drugs. *Clin. Microbiol. Revi.* **12(2):**286–287, 1999.
43. Landry, M. L., Stanat, S., Biron, K., et al. *A* standardized plaque reduction assay for determination of drug susceptibilities of cytomegalovirus clinical isolates. *Antimicrob. Agents Chemother*. **44(3):**688–692, 2000.
44. Chou, S., Marousek, G., Guentzel, S., et al. Evolution of mutations conferring multidrug resistance during prophylaxis and therapy for cytomegalovirus disease. *J. Infecti. Dis.* **176(3):**786–789, 1997.
45. Lurain, N. S., Bhorade, S. M., Pursell, K. J., et al. Analysis and characterization of antiviral drug-resistant cytomegalovirus isolates from solid organ transplant recipients. *J. Infecti. Dis.* **186(6):** 760–768, 2002.
46. Dankner, W. Scholl, M., D., Stanat, S. C., Martin,V., Sonke, R. L., and Spector, S. A., Rapid antiviral DNA-DNA hybridization assay for human cytomegalovirus. *J. Virol. Methods* **28(3):**293–298, 1990.
47. Chou, S., Erice, A., Jordan, M. C., et al. Analysis of the UL97 phosphotransferase coding sequence in clinical cytomegalovirus isolates and identification of mutations conferring ganciclovir resistance. *J. Infecti. Dis.* **171(3):**576–583, 1995.
48. McSharry, J. J., Lurain, N. S., Drusano, G. L., et al. Rapid ganciclovir susceptibility assay using flow cytometry for human cytomegalovirus clinical isolates. *Antimicrob. Agents Chemother.* **42(9):**2326–2331, 1998.
49. Chou, S., Guentzel, S., Michels, K. R., Miner, R. C., and Drew, W. L. Frequency of UL97 phosphotransferase mutations related to ganciclovir resistance in clinical cytomegalovirus isolates. *J. Infecti. Dis.* **172(1):**239–242, 1995.
50. Chou, S., Waldemer, R. H., Senters, A. E., et al. Cytomegalovirus UL97 phosphotransferase mutations that affect susceptibility to ganciclovir. *J. Infecti. Dis.* **185(2):**162–169, 2002.

51. Chou, S., Lurain, N. S., Thompson, K. D., Miner, R. C., and Drew, W. L. Viral DNA polymerase mutations associated with drug resistance in human cytomegalovirus. *J. Infect. Dis.* **188:**32–39, 2003.
52. Chou, S., Miner, R. C., and Drew, W. L. A deletion mutation in region V of the cytomegalovirus DNA polymerase sequence confers multidrug resistance. *J. Infect. Dis.* **182(6):**1765–1768, 2000.
53. Lurain, N. S., Weinberg, A., Crumpacker, C. S., and Chou, S. Sequencing of the cytomegalovirus UL97 gene for genotypic antiviral resistance testing. *Antimicrob. Agents Chemother.* **45:** 2775–2780, 2001.
54. Chou, S., Lurain, N. S., Weinberg, A., Cai, G. Y., Sharma, P. L., and Crumpacker, C. S. Interstrain variation in the human cytomegalovirus DNA polymerase sequence and its effect on genotypic diagnosis of antiviral drug resistance. Adult AIDS Clinical Trials Group CMV Laboratories. *Antimicrob. Agents Chemother.* **43(6):** 1500–1502, 1999.

IX Applications of Molecular Diagnostics to Identity-Based Testing

39 HLA Typing Using Molecular Methods

JOHN L. SCHMITZ

1. INTRODUCTION

Histocompatibility testing is a specialized area of clinical laboratory science with particular relevance to transplant programs. Although entailing several different specific functions, histocompatibility testing is most often equated with the determination of the human leukocyte antigen (HLA) phenotype or genotype of an individual (referred to as HLA typing). As discussed in this chapter, the approaches to HLA typing have evolved over the years. This evolution has been driven by the development of technologies, allowing a more detailed and accurate assessment of an individual's HLA type and the understanding that this more detailed information has great clinical value. The clinical applications of HLA typing most commonly inlcude (1) typing for renal transplantation as one component in the scheme to allocate deceased donor kidneys to individuals on the renal transplant waiting list because of the long-term benefit of using HLA-matched kidneys, (2) typing for hematopoeitic stem cell transplantation (HSC) as it is well known that the degree of HLA match between donor and recipient has a significant influence on the risk of serious adverse events associated with this process, and (3) typing as an aid in the diagnosis of HLA-associated diseases, the best known example being the association of the HLA-B27 type with ankylosing spondylitis.

The HLA phenotype of an individual consists of the array of HLA-specific proteins, encoded by genes within the major histocompatibility complex (MHC), displayed on the surface of most cells of the body. Although there are over 40 genes located in the HLA region, we will consider only those genes that encode the classical transplantation antigens (i.e., those that have been shown to influence the outcome of transplantation). These antigens include the MHC class I HLA-A, HLA-B and HLA-C molecules and the MHC class II HLA-DR, HLA-DQ and HLA-DP molecules. For decades, the class I HLA type has been defined immunologically using antisera obtained from individuals sensitized to HLA antigens via pregnancy, transfusion, or transplantation or with monoclonal antibodies. In this approach (Fig. 1), viable lymphocytes, to be HLA typed, are incubated with extensive panels of these antisera and complement. The panels of antibody reagents include sufficient numbers to detect most of the known HLA antigenic types. When a cell expresses a specific HLA antigen that is defined by a particular antiserum, the antibodies will bind and activate the complement, resulting in cell death that is detected by the addition of a viability dye. Using this approach, a number of specificities for the HLA-A, HLA-B, and HLA-C loci can be defined (Table 1). This high degree of polymorphism is the hallmark of the HLA system.

Polymorphism within the class II (HLA-DQ, HLA-DR, and HLA-DP) loci was originally defined by use of the mixed-lymphocyte culture. In this procedure, lymphocytes from two HLA class II disparate donors undergo blastogenesis when placed in culture together. Differences in class II (Dw) antigens stimulate this immune reaction. Cells homozygous for class II determinants, homozygous typing cells, have been used to define the Dw type of individuals. Eventually, antisera were defined and used to serologically identify an individual's HLA-DR and HLA-DQ type. HLA-DP has proven difficult to type serologically and molecular methods are employed to type this locus. Although less than the class I region, there are a number of different phenotypes defined by serologic means (Table 1).

As in many areas of clinical laboratory science, the application of molecular biologic techniques, has flourished in the HLA laboratory. Importantly, the application of polymerase chain reaction (PCR)-based HLA genotyping methods to histocompatibility testing made apparent the fact that the extent of HLA phenotypic polymorphism greatly underestimates the true degree of HLA polymorphism. Detailed analyses of HLA genes lead to the determination that several alleles could encode proteins, all recognized as a single serologic specificity. Immunologic studies demonstrated that these small differences, in some cases only a single amino acid difference between two HLA molecules, although not serologically distinguishable, could be recognized by the cellular arm of the immune system as foreign and lead to cellular immune reactivity. The clinical impact of this increased ability to distinguish variant alleles has become clear, particularly in the context of HSC transplantation. In addition to the more detailed definition of HLA polymorphism, genotypic analysis has other, more practical, advantages that have resulted in a transition from serologic to DNA-based typing in a large number of clinical histocompatibility laboratories.

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

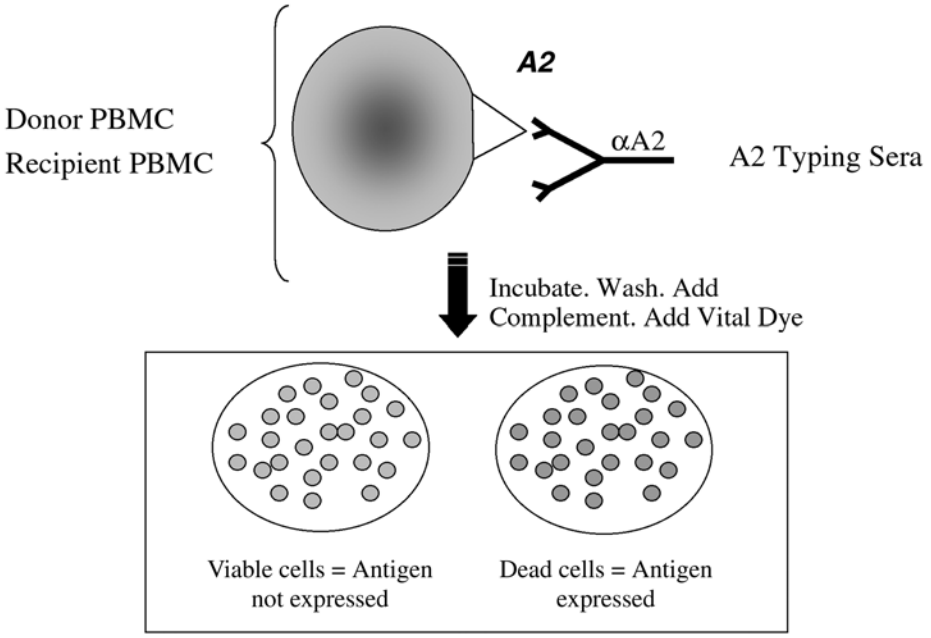


Fig. 1. Serologic HLA typing. Cells with an unknown HLA phenotype (e.g., donor and/or recipient lymphocytes) are incubated with panels of antisera (in this example and antiserum containing antibodies to the HLA-A2 antigen), each of which are specific for one or more HLA antigens at the class I or II loci. After an initial incubation, complement is added. If the antisera react with HLA antigens expressed by the cells, complement activation occurs and cell death ensues. The addition of a vital dye allows the observer to detect dead (orange) or live cells (green). A positive result (meaning cell death indicating recognition of the unknown HLA antigen by the typing serum) is indicated by greater than 20% cell death in the reaction. This method suffers from the fact that viable cells, typically lymphocytes obtained from blood, lymph nodes, or spleen, are needed, thus placing restrictions on sample handling. In addition, typing sera are difficult to obtain and prepare and are often polyspecific, which makes it necessary to use large numbers of antisera to be able to clearly define the presence of an HLA antigen. (Figure appears in color in accompaning CD ROM.)

Table 1
Serologically Defined HLA Specificities

HLA-A		*HLA-B*		*HLA-C*	*HLA-DR*	*HLA-DQ*
A1	B5	B42	B65(14)	Cw1	DR1	DQ1
A2	B7	B44(12)	B67	Cw2	DR103	DQ2
A203	B703	B45(12)	B70	Cw3	DR2	DQ3
A210	B8	B46	B71(70)	Cw4	DR3	DQ4
A3	B12	B47	B72(70)	Cw5	DR4	DQ5(1)
A11	B13	B48	B73	Cw6	DR5	DQ6(1)
A23(9)	B14	B49(21)	B75(15)	Cw7	DR6	DQ7(3)
A24(9)	B15	B50(21)	B76(15)	Cw8	DR7	DQ8(3)
A2403	B16	B51(5)	B77(15)	Cw9(w3)	DR8	DQ9(3)
A25(10)	B17	B5102	B78	Cw10(w3)	DR9	
A26(10)	B18	B5103	B81		DR10	
A28	B21	B52(5)	Bw4		DR11(5)	
A29(19)	B22	B53	Bw6		DR12(5)	
A30(19)	B27	B54(22)			DR13(6)	
A31(19)	B2708	B55(22)			DR14(6)	
A32(19)	B35	B56(22)			DR1403	
A33(19)	B37	B57(17)			DR1404	
A34(10)	B38(16)	B58(17)			DR15(2)	
A36	B39(16)	B59			DR16(2)	
A43	B3901	B60(40)			DR17(3)	
A66(10)	B3902	B61(40)			DR18(3)	
A68(28)	B40	B62(15)			DR51	
A69(28)	B4005	B63(15)			DR52	
A74(19)	B41	B64(14)			DR53	
A80						

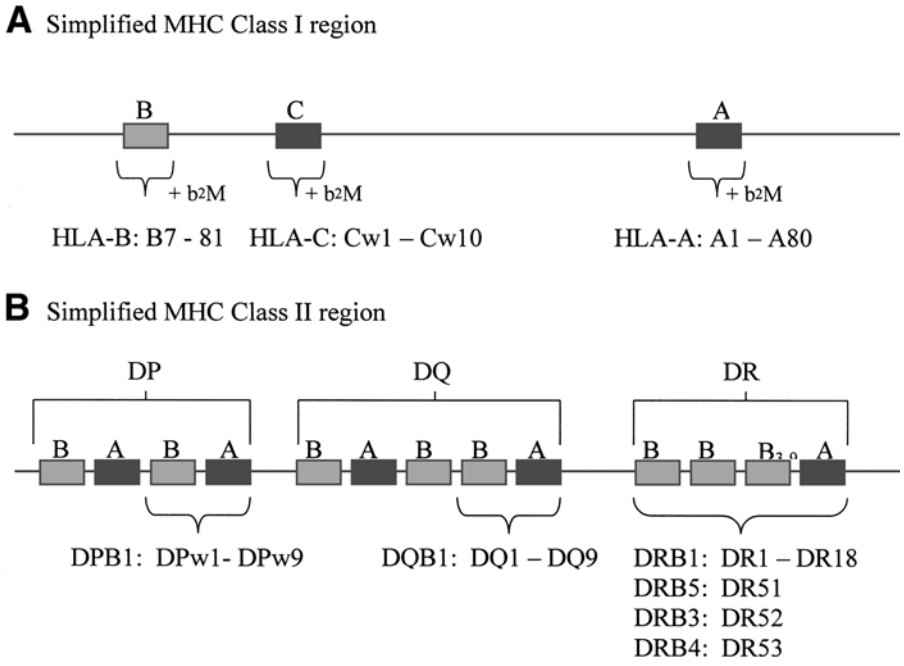


Fig. 2. A simplified physical map of the MHC. **(A)** A representation of the organization of the class I region with only the classical transplant antigens shown. Also indicated are the protein products of these loci. **(B)** A representation of the MHC class II region. The class II serologic specificities (shown) endcoded in this region are the products of an A and B gene as indicated. Serologic specificities correspond to polylmorphisms endcoded by the B-chain genes at each of these loci.

2. STRUCTURE–FUNCTION OF HLA MOLECULES

Human leukocyte antigens are glycoproteins expressed on the cell surface that play a critical role in the adaptive immune response of presenting antigenic peptides to T-lymphocytes. The class I molecules present peptides originating in the cell cytoplasm for recognition by peptide-specific CD8-positive T-lymphocytes. The recognition by and subsequent activation of the CD8 T-lymphocyte could then lead to cytolysis of the infected cell. The class II molecules present peptide antigens, derived from extracellular sources, to peptide-specific CD4-positive T-lymphocytes. CD4 T-cell activation allows them to provide help for the production of antibody by B-lymphocytes or for CD8 T-lymphocyte-mediated cytotoxicity. Because of the enormous array of pathogens encountered by humans, and the sequence-specific restriction of peptide binding to HLA molecules, the HLA system has evolved to contain several class I and II loci as well as multiple alleles within each locus. This high degree of polymorphism is beneficial, as it allows the human population to respond immunologically to a wide variety of foreign antigens and thus enhance survival. Although this role in protective immune responses is critical, it is problematic for solid organ and bone marrow transplantation because class I and II HLA proteins are highly immunogenic and are thus a "major" stimulus for graft rejection and graft-vs-host disease. Fortunately, the ability of laboratories to identify a person's array of HLA antigens or alleles facilitates the identification of suitable donor/recipient combinations for transplantation.

3. GENETICS OF THE MAJOR HISTOCOMPATIBILITY COMPLEX

Human leukocyte antigens are encoded in the MHC on the short arm of chromosome 6. There are three groups of MHC genes, clustered in the the class I, II, and III regions (Fig. 2). The class II region is centromeric and encodes, among many other genes, the HLA-DP, HLA-DQ, and HLA-DR locus genes. The class III region is telomeric to the class II region and, although encoding many genes, does not contain genes for the classical transplantation antigens. The class I region is telomeric and encodes the HLA-B, HLA-C and HLA-A loci. The array of class I and II genes on a chromsome is termed a haplotype. Haplotypes are inherited in a Mendelian fashion.

The class I region genes, HLA-A, HLA-B, and HLA-C each encode a single polypeptide chain that is expressed on the cell surface in association with β_2-microglobulin. The genes are codominantly expressed so that individuals have two HLA-A, two HLA-B, and two HLA-C antigens expressed on the cell surface. There are null alleles present in the MHC region whose products are not expressed on the cell surface; however, they are infrequent. The class I genes all have a similar structure consisting of multiple exons, the second, third and fourth of which encode the extracellular domains of the HLA molecule. The extracellular α1 and α2 domains of the protein that comprise the peptide-binding groove of the molecule are encoded by the second and third exons. These domains contain regions of hypervariability defining the various alleles. The polymorphisms within these second and third exons are the targets of DNA-based typing approaches discussed below.

The class II molecules at each of the three loci are heterodimers whose codominantly expressed proteins are encoded by an A (HLA-DRA, HLA-DQA, and HLA-DPA) and B (HLA-DRB, HLA-DQB, and HLA-DPB) gene. Similar to class I heavy chains, the class II α- and β-chains also have a domain structure in which polymorphisms are primarily restricted to the membrane distal domains that together comprise the antigen-binding

Table 2
HLA Allele Numbers by Locus

Locus	*No. of Alleles*
HLA-A	349
HLA-B	626
HLA-C	182
HLA-DRB1	390
HLA-DRB3	41
HLA-DRB4	13
HLA-DRB5	18
HLA-DQB1	60
HLA-DPA1	22
HLA-DPB1	116

Source: Adapted from www.anthonynolan.org.uk/HIG/.

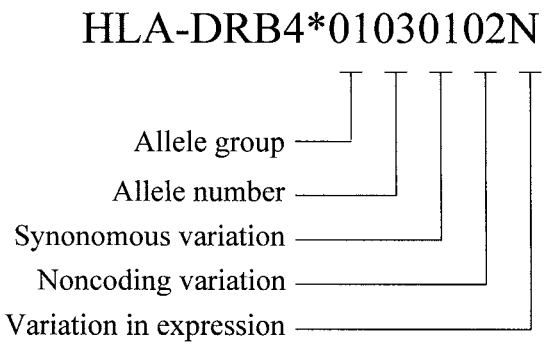


Fig. 3. An example of HLA nomenclature as defined by DNA-based typing. The gene being typed is given first followed by an "*". The first two numerals of the designation correspond to the allele group encoded by the gene. This might be equivalent to the serologic specificity of the antigen encoded by that gene. The next two numerals indicate the specific allele number of the gene. Two additional numerals might follow, indicating the presence of a synonymous polymorphism. The final two numerals indicate polymorphisms in noncoding regions of the gene. Finally, a letter code might be present at the end of the allele name giving information about the expression of the gene: (N) null = not expressed; (C) cytoplasmic = present in the cytoplasm but not cell surface; (S) secreted molecule; (A) aberrant expression = the expression of the antigen is unclear.

groove. For the B genes, polymorphisms are concentrated within the second exon that encodes the $\alpha 1$ domain of the molecule. Polymorphism within the A chain genes varies by locus with the DRA gene being nonpolymorphic and the DQA and DPA genes polymorphic. Polymorphisms detected by HLA typing laboratories are based on the B-chain genes for each of the three class II loci. At the DR locus, the DRB1 gene encodes the DR1 through 18 serologically defined specificities. The DRB3, 4, and 5 genes, encoding the DR52, DR53, and DR51 antigens, respectively, are not present in all individuals—only in certain haplotypes. The DRB5 gene is present in haplotypes containing the DRB1*15 and DRB1*16 allele groups. The DRB4 gene is present in haplotypes containing the DRB1*04, DRB1*07, and DRB1*09 allele groups. The DRB3 gene is present in haplotypes containing the DRB1*03, DRB1*11, DRB1*12, DRB1*13, and DRB1*14, allele groups.

As stated earlier, the hallmark of the HLA genes is their high degree of polymorphism. Not only are there three loci within each of the class I and II regions, there are many alleles at each of the loci. As shown in Table 2, the number of alleles at each locus far outweighs the number of serologically defined specificities. It is important to remember, as discussed below, that the cellular immune system can resolve these differences (i.e., mount an immune response) that might be as minimal as a single amino acid substitution.

4. HLA NOMENCLATURE

Human leukocyte antigen typing had previously been carried out using serologic methods. The nomenclature employed for serologic typing includes a specification of the HLA locus being typed followed by a one- or two-digit number indicating the specificity (see Table 1). HLA alleles are named in a similar fashion but in a way to accommodate a much larger number of allelic determinants than serologic specificities (Fig. 3). The molecular nomenclature again begins with a designation of the HLA gene being typed (e.g., HLA-DRB1) followed by an asterisk that separates the gene name from a two-digit number that defines the allele group for that gene and in many cases corresponds to the serologic group encoded by that allele group. This level of HLA typing is termed "low resolution" (equivalent to the serologic specificity or allele group). If one is typing a specific allele group at a locus that contains a large number of allelic variants and can narrow the identification to a few potential alleles in the group, equivalent in some cases to identification of a serologic split, one has performed an "intermediate resolution" typing. If allele-level typing (i.e., high-resolution typing) is performed, the two-digit allele group number is followed by two additional numerals that identify the specific allele in the order they were identified (e.g., HLA-A*0301). With the identification of the extreme variability of the HLA region came the identification of silent or synonymous substitutions in coding regions of some alleles. These silent polymorphisms are distinguished by a fifth and sixth numeral in the allele number (e.g., HLA-A*020101 and HLA-A*020102 specify two alleles that differ in nucleic acid sequence but not in their protein product). Allele designations might have an additional two digits that indicate polymorphisms identified in noncoding regions of the gene (introns or 5′ or 3′ regions). Finally, some allele designations end in the letter N, S, C, or A to indicate a null allele (not expressed on the cell surface), a secreted molecule, a cytoplasmic molecule, or a molecule with aberrant expression (i.e., there is doubt as to the expression of the molecule), respectively. With the ever-increasing number of HLA alleles being identified, it has become apparent that the current system will not accommodate the number of alleles that will be identified in certain allele groups. As such, a modification to the nomenclature system has been developed to accommodate allele groups with more than 99 allelic variants. For allele groups in which the number of alleles identified exceeds 99, an additional number series will be established.

5. MOLECULAR METHODS FOR HLA TYPING

As aluded to earlier, the ability to identify allelic variants of HLA encoded genes has important implications. Although there are numerous methods one can use for DNA-based HLA

typing, they can be categorized into two broad groups: those that are based on detection of HLA allele or allele group sequences via primers, probes, or sequencing reactions and those that rely on the conformation of amplified HLA genes and their migration in electrophoretic gels. Several of the methods described in Section 6 are available as commercial kits from vendors.

6. METHODS BASED UPON SEQUENCE DETECTION

6.1. POLYMERASE CHAIN REACTION WITH SEQUENCE-SPECIFIC PRIMERS One of the most common methods employed for molecular HLA typing is PCR–sequence-specific primers (PCR-SSP). A major reason for this tests popularity lies in its simplicity (Fig. 4A). The basis of this approach is the use of multiple PCR reactions employing a panel of primer pairs that amplify specific alleles (high resolution) or allele groups (low/intermediate resolution). These primers target the second and third exons for class I and the second exon of class II genes. The amplification products are electrophoresed in an agarose gel and stained with ethidium bromide for visualization of bands. If a specific allele or allele group sequence complementary to the forward and reverse primers is present in the DNA sample, an amplification product of a specific size will be present on the gel. No target-specific amplification indicates the lack of an allele or allele group sequence that is complementary to the primers. The absence of amplification product is the result of the presence of mismatches at the 3′ end of one or both primers. Because *Taq* polymerase does not have proof-reading capabilities, these mismatched 3′ ends are not repaired during amplification.

The number of primer pairs that must be used to generate an HLA type depends on the locus being tested and the level of resolution required. An example of an agarose gel of PCR-SSP amplification products is shown in Fig. 4b.

Because this system of typing depends on the presence or absence of an amplified product, an internal control is included in each amplification reaction to ensure that lack of amplification is the result of the absence of a specific target sequence and not of an amplification failure. Internal controls typically consist of a primer pair that amplifies a ubiquitous housekeeping gene such as β-globin. The product of the internal control amplification is of a distinct size so as not to obscure the amplified HLA product.

The PCR-SSP system is a simple, rapid method of HLA typing. It can be used to obtain low-, intermediate-, and high-resolution HLA types in a short time. Its major limitation is the large number of primer pairs needed to obtain higher-resolution typing. This limitation is particularly noteable for higher-volume laboratories. Other methods are available, based on probe hybridization reactions that are amenable to both low- and high-volume applications.

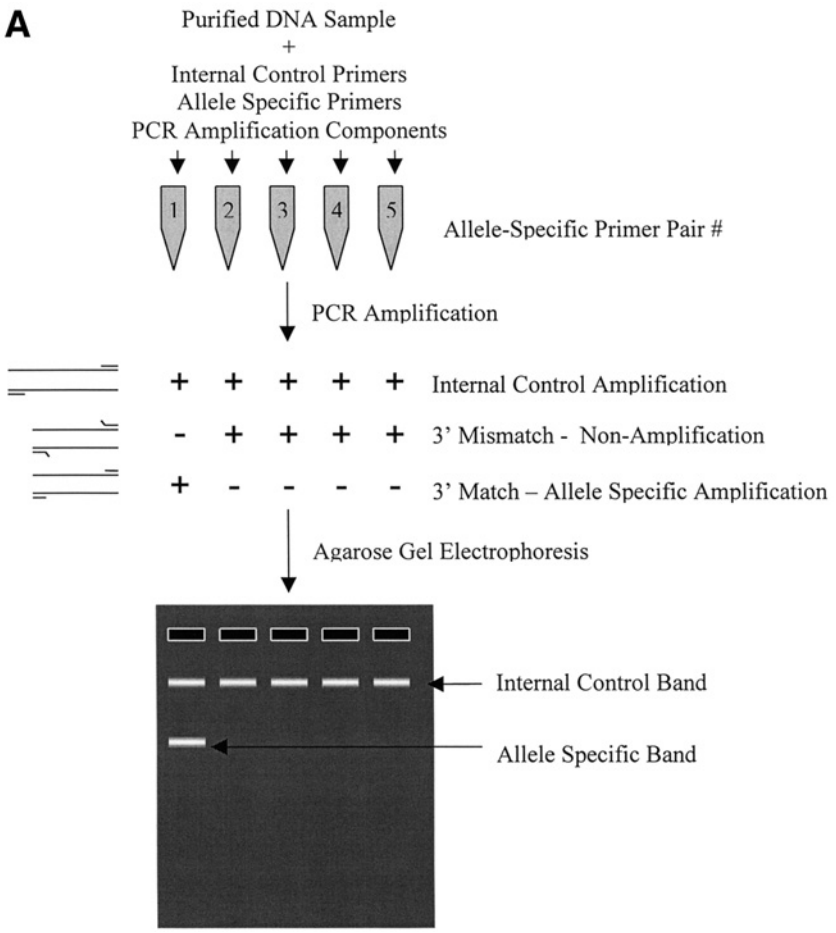
6.2. SEQUENCE-SPECIFIC OLIGONUCLEOTIDE PROBE HYBRIDIZATION Sequence-specific oligonucleotide probe (SSOP) hybridization is an additional method for HLA typing that is in common use and is being adapted to multiple formats so as to accommodate varying workloads of clinical laboratories. Instead of using multiple primer pairs targeted at polymorphisms unique to specific alleles or groups of alleles, SSOP employs locus-specific (generic) primer pairs followed by hybridization of the amplified DNA with panels of allele-specific or allele-group-specific oligonucleotide probes. The probes are constructed with an enzymatic or fluorescent label to allow the detection of hybrids. As for PCR-SSP primers, the oligonucleotide probes in SSOP assays target polymorphisms in the second and third exons of class I genes and the second exon of class II genes. This assay can be performed in the forward (sample DNA bound to solid phase; particularly amenable to high-volume laboratories) or reverse (probes bound to solid phase; suitable for lower-volume laboratories) direction and on a number of different solid support matrices.

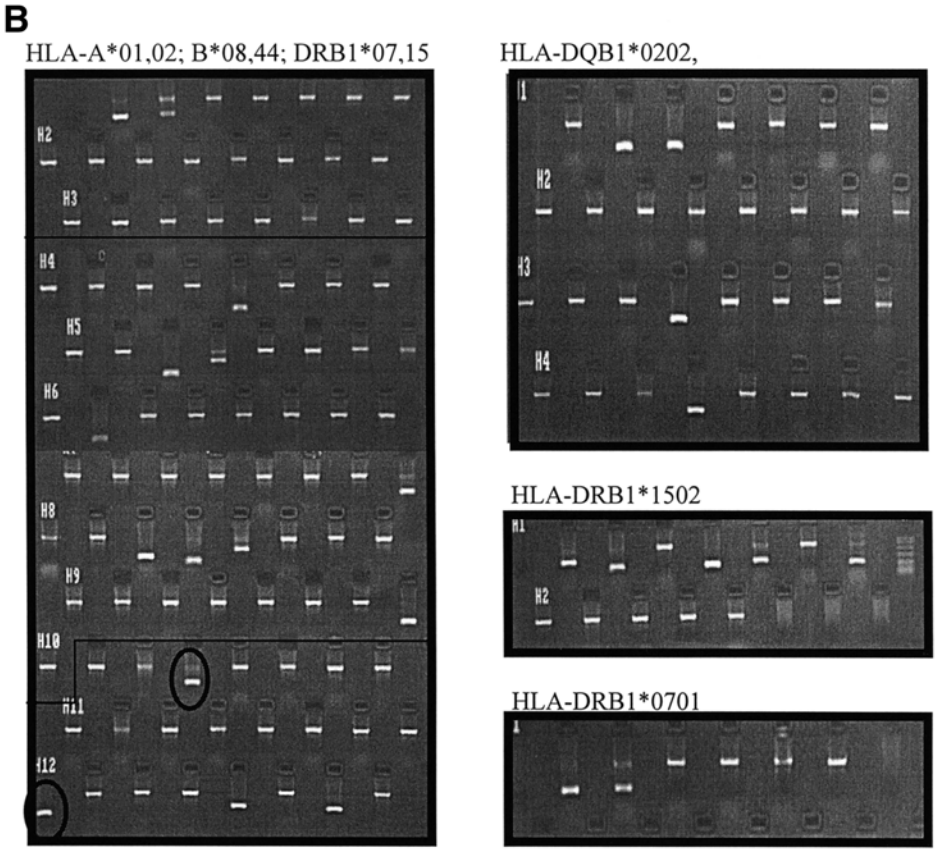
As with PCR-SSP, there are a number of commercial vendors that sell SSOP HLA typing kits. Although they vary in the specific primers, probes, solid phase, and reporters used, they follow the same basic principle of a generic locus-specific amplification followed by hybridization with a series of probes to determine the HLA allele group of a DNA sample. To obtain intermediate-resolution typing, some kits require additional amplifications to unambiguously assign the intermediate-resolution type (i.e., serologic group). Additional locus-specific probes are required to obtain allele-level typings.

The SSOP typing can be performed in several different formats. Microplate-based methods have probes bound to the wells of a microtiter plate with hybridizations and color development taking place in the wells of the plate. With the application of automated plate handling systems, laboratories can have a highly automated approach to HLA typing. Another variation involves the application of Luminex®-based technology. Panels of beads, each with a unique fluorescent signature and bound probe, are hybridized with locus-specific amplified DNA. Hybrids are detected by virtue of a fluorescent tag on the amplified product using a "mini" flow cytometer. The identity of the allele or allele group is determined by the specific bead to which the amplified, fluorescently tagged DNA hybridizes. Finally, line-probe-based assays (oligonucleotide probes bound to membrane strips) with automated processors, scanners, and interpretive software are available to enhance laboratory throughput.

A limitation inherent in both PCR-SSP and SSOP methods is that they interrogate only a small segment of the DNA sequence of the gene. As such, polymorphism that are present outside of these areas are missed. DNA sequencing is the only method in common use that provides a complete assessment of the presence of polymorphisms.

6.3. SEQUENCE-BASED HLA TYPING DNA sequencing is an established technology in the research laboratory that, with developments in instrumentation and technology, is amenable to implementation in the clinical laboratory. Standard approaches employing automated sequencers have been applied to HLA typing and commercial reagent kits are available to laboratories. The standard approach to sequencing involves isolation of DNA, locus-specific PCR, sequencing reactions, and analysis with a sequencing instrument and software. Kits are available to type all of the classical transplantation loci including HLA-A, HLA-B, HLA-C, HLA-DR,


A
Purified DNA Sample
+
Internal Control Primers
Allele Specific Primers
PCR Amplification Components
1
2
3
4
5
Allele-Specific Primer Pair #
PCR Amplification
+ + + + +
Internal Control Amplification
- + + + +
3' Mismatch - Non-Amplification
+ - - - -
3' Match – Allele Specific Amplification
Agarose Gel Electrophoresis
Internal Control Band
Allele Specific Band


B
HLA-A*01,02; B*08,44; DRB1*07,15
HLA-DQB1*0202,
HLA-DRB1*1502
HLA-DRB1*0701

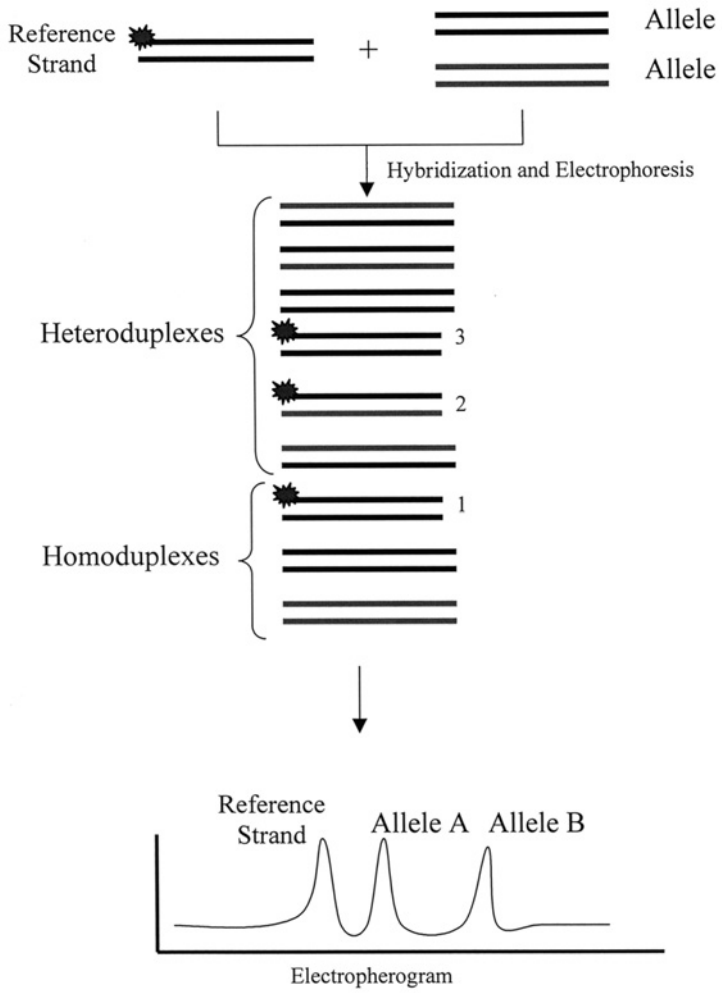


Fig. 5. A model of RSCA-based HLA typing. Purified DNA (containing hypothetical alleles A and B) is denatured and rehybridized with a reference DNA containing one fluorescently labeled strand. Homoduplexes and heteroduplexes are formed upon hybridization, as indicated. This mixture is electrophoresed in an automated DNA sequencing device and yields and electorpherogram demonstrating peak corresponding to reference-strand homoduplexes as well as heteroduplexs corresponding to alleles A and B. Not shown for clarity are two size standard calibration peaks present in each lane. The migration rate of the heteroduplexes is compared to a database of known alleles to determine the HLA type.

HLA-DQ, and HLA-DP. This approach to HLA typing, with current instrumentation and software provides high throughput with allele-level resolution.

Sequencing is required for the definition of new HLA alleles because it is not limited to interrogation of a limited sequence of nucleic acids, as PCR-SSP primers and SSOP probes are. As with other methods, however, there are ambiguous combinations of alleles in heterozygotes that cannot be resolved because of the inability to establish cis or trans linkage of the polymorphisms.

7. METHODS BASED ON DETERMINATION OF CONFORMATION

7.1. SINGLE-STRANDED CONFORMATIONAL POLYMORPHISM Single-stranded conformational polymorphism has been applied to the detection of mutations in a variety of genetic diseases. The approach has also been applied to HLA typing. This method involves locus-specific PCR amplification followed by denaturation of the PCR product. The denatured products are then electrophoresed in a

Fig. 4. (A) Model of a PCR-SSP typing assay. Sample DNA is amplified with a panel of primers specific for alleles within a hypothetical group and an internal control primer pair. The amplification products are then electrophoresed in an agarose gel. The first primer pair in the panel (tube 1) amplifies the first allele of the group. This in reflected in the gel by the presence of an internal control as well as specific amplification product of a smaller size. The remaining lanes of the agarose gel show only the internal control indicating that successful amplification has occurred but that no DNA containing the complementary polymorphisms for the other specific primers (tubes 2–5) is present. **(B)** An example of PCR-SSP typing. The 96-lane gel shows amplification products from and HLA-A, HLA-B and HLA-DR typing. Lanes with positive HLA-specific amplification products show an band of smaller size compared to the internal control. The top three rows of this gel provide a low-resolution typing of the HLA-A locus, The next six rows provide a low resolution typing of the HLA-B locus. The final three rows provide a low-resolution typing of the HLA-DRB1 locus. The panel in the upper right quadrant is a high-resolution HLA-DQB1 PCR-SSP gel. The smaller panels to the right are high-resolution typings of the HLA-DRB1*15 and HLA-DRB1*07 allele groups.

nondenaturing polyacrylamide gel. During electrophoresis, the single-stranded DNA molecules assume unique conformations dependent on their sequence and the length. SSCP has proven to be a sensitive method for the detection of sequence variation capable of detecting even single-base substitutions. Coupled with its simplicity and performance with no specialized equipment, it is an attractive approach to the determination of HLA types. However, this method is limited by the fact that the gels can be complex and difficult to interpret.

7.2. HETERODUPLEX ANALYSIS As with SSCP, heteroduplex analysis (HA) relies upon the conformation of DNA during electorphoresis. However, HA involves the locus-specific amplification of DNA, denaturation, and then reannealing of strands. Homoduplexes and heteroduplexes will form during the renaturation process. These species of renatured DNA have distinct migratory properties in the electrophoretic gel because of the presence of "bubbles" or "buldges" in the heteroduplexes reflecting mismatches in base-pairing. As with SSCP, this is a relatively simple approach to HLA typing but also suffers from the fact that gels might be difficult to interpret.

7.3. REFERENCE STRAND CONFORMATION ANALYSIS Reference strand conformation analysis (RSCA) is a more recently developed variation of HA that can provide high-resolution typing of HLA class I and II alleles. This method relies on the conformational properties of amplified target DNA and locus-specific fluorochrome-labeled reference DNA hybrids. These hybrids will have distinct conformations, leading to variations in migration rates that can be detected in an automated DNA sequencer. The migration rates of an unknown sample can be compared to a database of known alleles to identify the specific allele. Alternatively, this approach can be used for matching donor and recipient based solely on migration rates without knowledge of allele identity.

The process of RSCA (Fig. 5) begins with a locus-specific amplification of target DNA. Amplified DNA is then hybridized with one or more fluoresecently labeled, locus-specific reference DNA strands. The reference strand will bind to the target DNA sense strand in a conformation unique to the specific allele. Both heteroduplexes and homoduplexes will be formed. However, only hybrids that have incorporated a labeled reference strand will be detected in an automated sequencing instrument. Heteroduplexes will have stretches of complementary binding along with mismatched segments that result in the formation of bubbles at various locations along the hybrids. Because each HLA allele will have a unique conformation in the heteroduplex, their migration rates in a polyacrylamide gel will differ. The electropherogram of a sample will demonstrate several peaks. The fastest and slowest migrating peaks are size standards present in the samples applied to the gel that serve as calibrators to assure lane-to-lane comparability. Reference-strand homoduplexes will be the next observed peaks followed by one or two peaks that represent target DNA heteroduplexes. Homozygous samples will demonstrate one peak and heterozygous samples will demonstrate two peaks. There can be occasions when the target DNA strand is homologous to the reference-strand. In this case, only one peak in addition to the size standards and reference-strand homoduplex will be seen. However, because the commercial version of this assay employs two or more reference strands, the second reference-strand hybridization will typically demonstrate two target heteroduplexes.

Reference strand conformation analysis is a reproducible assay that is fairly simple to perform with the appropriate equipment and has a high throughput. It can distinguish alleles differing by only a single nucleotide. It can also resolve ambiguous combinations of HLA alleles found by other methods. Although there are instances where two alleles might have the same mobility with a particular reference strand, the use of two or more reference strands will usually allow resolution of the alleles. Although the commercially available version of this method can resolve most of the more common HLA alleles, the mobility patterns for all alleles have not been determined and, thus, one might not be able to identify rare alleles.

8. CLINICAL IMPACT OF MOLECULAR-BASED HLA TYPING

The impact of molecular methods of HLA typing is manifested in several ways. The most significant impact has been achieved in the field of HSC transplantation. It has been known for decades that successful HSC transplantation is dependent on matching donor and recipient HLA antigens to reduce the risk for graft loss and graft-vs-host disease. Serologic methods (or low-resolution molecular methods) of HLA typing are sufficient when inheritance of HLA haplotypes could be confirmed in related recipients and donors. However, most individuals in need of an HSC transplant do not have an HLA-identical sibling, thus alternative donors, matched unrelated and HLA mismatched, have been used. Because serologic methods of HLA typing are insufficient to identify the extent of HLA compatibility between unrelated individuals, DNA-based methods became a necessity.

Numerous studies have documented the fact that allele-level matching is clinically important in HSC transplantation. Mismatches between donor and recipient can have two undesirable consequences: graft failure and graft-vs-host disease. Immunologic causes of graft failure stem from the recognition of mismatched donor HLA antigens by the recipients immune system. Presensitization to donor HLA antigens might increase the risk of graft failure. Graft-vs-host disease is a potentially fatal complication arising from the recognition of mismatched recipient HLA antigens by immunocompetent donor cells. Manifestations in the skin, gastrointestinal tract, and liver can be very serious. As such, recipients are routinely immunosuppressed for extended periods of time after transplantation to reduce the risk of developing this complication.

Molecular-based DNA typing methods play a critical role in reducing the risk of developing these complications by identifying HLA genes to the allele level. Two recent, large studies have clearly demonstrated that allele-level matching is critical to reduce the risk of these complications. The roles of individual loci, both class I and II, have been studied for their relative contribution to the development of these complications. It is still unclear as to the relative significance of mismatches at the class I and II loci. However, it is clear that multiple mismatches are to be avoided if possible.

Molecular-based typing methods are also used in solid organ transplant programs. However, the benefit of these methods lies not in their ability to identify donor and recipient mismatches at the allele level. Rather, their most important benefit is their accuracy compared to serologic-based methods. Even though HLA matching has been clearly demonstrated to improve graft and patient survival in renal and cardiac transplantation, the limited donor pool makes it impossible to use allele-level matching in any allocation system. The use of DNA-based typing methods has been shown to improve the accuracy of HLA typing. As such, there is an inherent benefit to the use of these methods to improve outcomes. In addition to accuracy, other advantages of DNA-based methods of HLA typing include the fact that primers and probes are easily prepared or purchased and consistent as opposed to HLA typing sera, which might be in limited supply and often are polyspecific. Furthermore, serologic typing must employ viable lymphocytes, as the readout is cell death. DNA typing can be carried out on any source of good quality DNA irrespective of cell viability. Finally, methods such as PCR-SSP can be performed in the same time frame as serologic methods, thus making them applicable for cadaveric donor typing, which must be completed as expeditiously as possible.

9. CONCLUSIONS

The application of molecular biologic techniques has revolutionized the HLA laboratory. Using these methods, laboratories can provide rapid and accurate HLA typings that have clear clinical benefit. The variety of methods available allows a laboratory to employ the most appropriate method for their experience and workload. The advantages of DNA-based methods and the variations in format make this technology broadly applicable to a variety of HLA typing laboratories.

SUGGESTED READING

Abbas, A. K. and Lichtman, A. H. The major histocompatibility complex. In: Abbas, A. K., Lichtman, A. H., (eds.). *Cellular and Molecular Immunology,* 5th ed. Philadelphia: Saunders, 2003; pp. 65–80.

Arguello, J. R. and Madrigal, J. A. HLA typing by reference strand mediated conformation analysis (RSCA). *Rev. Immunogenet.* **1:**209–219, 1999.

Cros, P. and Allibert, P. Oligonucleotide genotyping of HLA polymorphism on microtiter plates. *Lancet* **340:**870–873, 1992.

Erlich, H. A. Polymerase chain reaction-based methods of HLA typing. In: Lechler, R., Warrens, A. (eds.). *HLA in Health and Disease,* 2nd ed. San Diego: Academic Press, 2000, pp. 451–461.

Flomenberg, N., Baxter-Lowe, L., Confer, D., et al. Impact of HLA class I and II high-resolution matching on outcomes of unrelated donor bone marrow transplantation: HLA-C mismatching is associated with strong adverse effect on transplantation outcome. *Blood* **104:**1923–1930, 2004.

Hoshino, S., Kimura, A., Fukuda, Y., Dohi, K., and Sasazuki, T. Polymerase chain reaction-single-strand conformation polymorphism analysis of polymorphism in DPA1 and DPB1 genes: a simple, economical, and rapid method for histocompatibility testing. *Human Immunol.* **33:**98–107, 1992.

Hui, K. M. and Bidwell, J. L. (eds.). *Handbook of HLA Typing Techniques.* CRC Press, 1993.

Hurley, C. K., Tang, T., Ng, J., Steiner, N., and Hartzman, R. J. HLA typing by molecular methods. In: Rose, N. R., Hamilton, R. G., and Detrick, B. (eds.). *Manual of Clinical Laboratory Immunology,* 6th ed., Washington DC: ASM Press, 2002, pp. 1134–1151.

Lechler, R. I., Simpson, E., and Bach, F. H. Major and minor histocompatibility antigens: an introduction. In: Bach, F. H. and Auchincloss, H., (eds.). *Transplantation Immunology,* New York: Wiley-Liss, 1995, pp. 1–34.

McGinnis, M. D., Conrad, M. P., and Bouwens, A. G. M. Automated, solid phase sequencing of DRB region genes using T7 sequencing chemistry and dye-labeled primers. *Tissue Antigens* **46:**173–179, 1995.

Mytilineos, J., Lempert M., Scherer, S., Schwarz, V., and Opelz, G. Comparison of serological and DNA PCR-SSP typing results for HLA-A and HLA-B in 421 black individuals: a Collaborative Transplant Study report. *Hum. Immunol.* **59:**512–517, 1998.

Olerup, O. and Zetterquist, H. HLA-DR typing by PCR amplification with sequence specific primers (PCR-SSP) in 2 hours: an alternative to serological DR typing in clinical practice including donor-recipient matching in cadaveric transplantation. *Tissue Antigens* **39:**225–235, 1992.

Sasazuki, T., Juji, T., Morishima, Y., et al. Effect of matching of class I HLA alleles on clinical outcome after transplantation of hematopoeitic stem cells from an unrelated donor. Japan Marrow Donor Program. *N. Engl. J. Med.* **339:**1177–1185, 1998.

Summer, C., Morling, F., Taylor, M., Yin, J. L., and Stevens, R. Donor-recipient HLA class I bone marrow transplant matching by multilocus heteroduplex analysis. *Transplantation* **58:**628–629, 1994.

40 Molecular Analysis for Forensic Casework and Parentage Testing

TRACEY DAWSON CRUZ

1. INTRODUCTION

Disciplines in forensic science have developed alongside and continue to be dependent on technological advances in the corresponding basic scientific disciplines. This is true with forensic medicine, toxicology, and chemistry but has become most evident with the advent of PCR (polymerase chain reaction) and molecular biology, a field that has rapidly made way for powerful DNA-based methodologies that can be easily used for human identification. Applications are wide ranging, from paternity/maternity establishment to determination of violent criminal offenders through forensic casework. The current molecular techniques employed in identity testing can be used to generate genetic profiles that are so rare (<1 in 360 billion) that some laboratories will state that identity has been demonstrated and an analyst can confidently report that a biological specimen originated *only* from a specific individual or his/her identical twin.

Over the last decade, the public has become more aware of the power of DNA typing. Several infamous identity cases have been covered extensively in the media, including the murder trial of O. J. Simpson, the President Clinton–Monica Lewinsky blue dress scandal, the identification of the remains of the tomb of the unknown soldier, the identification of the Romanoff family remains, and the identification of slave-born descendents of the third president of the United States, President Thomas Jefferson. Most recently, DNA identification techniques have been brought to the forefront because of the tremendous task of finding and identifying remains of the victims of the September 11, 2001 terrorist attacks. After the attacks, more than 20,000 total biological samples were recovered combined from the rubble of the World Trade Center, the soil at the site of the United Flight 93 crash in Somerset, PA, and from the American Airlines Flight 77 Pentagon crash site. Biological material recovered from the scenes consisted primarily of bone, teeth, and small samples of soft tissue, which ranged from fresh, to gangrenous, to carbonized. However, reference samples brought from family members included bloodstains, toothbrushes, hair, clothing items, and razors. In the case of this mass disaster, the role of DNA should not be understated—as of December 2002, of the total number of victim identifications made, approx 38% were made exclusively with DNA evidence. Additionally, an approx 40% of the identifications made were through the combined use of DNA along with a more traditional identification method (dental records, personal identifiers, forensic anthropology) *(1)*. Many of these identifications were made from minute amounts of charred, highly degraded biological samples that might have otherwise not been properly identified and returned to the families.

Whereas the above-mentioned cases have attracted widespread attention, it is important to note that they represent only a very small fraction of the tens of thousands of DNA-based identity cases that are conducted each year by both private commercial laboratories and publicly funded crime labs. In fact, in 2001 alone there were more than 310,000 cases representing nearly 1 million samples submitted for parentage testing to accredited laboratories in the United States *(2)*. The number of paternity cases tested in the United States has increased each year by approx 3% and has increased by more than 45% over the past 10 yr *(2)*. DNA technologies have expanded beyond the traditional identification cases and now include cases that involve identification of domestic animals *(3)*, wildlife species *(4)*, plant species *(5)*, and human identification from maggots found on decomposing bodies *(6)*. As the public becomes more and more fascinated with forensic science and DNA technologies, testing laboratories will continue to experience a tremendous increase in demand. In order for forensic scientists to accommodate the increasing need, more robust, more accurate, and higher-throughput automated methods will need to be developed, validated, and implemented. These techniques can help to improve existing protocols and ultimately will help to decrease backlogs that exist in crime laboratories across the country.

1.1. DNA TYPING: AN HISTORICAL OVERVIEW

DNA-based human identification began in the mid-1980s with the discovery of highly polymorphic genetic loci, VNTRs

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

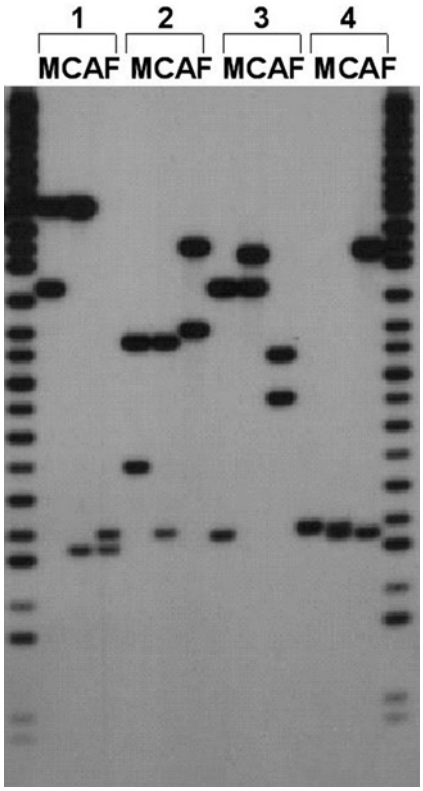


Fig. 1. RFLP autorad results from paternity casework. DNA samples from four traditional paternity trios (M = mother, C = child, AF = alleged father) were digested with *Hae*III, run on gel, and transferred to a nylon membrane via Southern blotting. The membrane was hybridized to a radioactively labeled probe (^{32}P) for EFD52 (a VNTR locus). The resulting autoradiograph is shown. The first and last lanes represent molecular-weight markers. Paternity results for this locus are indicated as follows: case 1, AF included; case 2, AF excluded; case 3, AF excluded; case 4, AF included. (Courtesy of Fairfax Identity Laboratories, a division of Commonwealth Biotechnologies, Inc.)

(variable number of tandem repeats), by English scientist Alec Jeffreys *(7,8)*. VNTR repeats ("minisatellites") range in length from 10 to 80 bp and are found scattered throughout most chromosomes. These loci were ideal for identity testing because the number of repeats found in a specific VNTR locus is highly variable between unrelated individuals within a population. Thus, the first method of DNA analysis applied in forensic identification was RFLP (restriction fragment length polymorphism) analysis of these VNTR loci. With this procedure, DNA is digested with a restriction endonuclease (typically *Hae*III). As shown in Fig. 1, the resulting fragments are separated by gel electrophoresis, transferred to a nylon membrane via Southern blot techniques, and subsequently detected by hybridrization to either a radioactive or chemiluminescent VNTR probe *(9,10)*. Traditionally, because of the extremely high discriminatory power of these hypervariable loci, only three to five probes are analyzed for a typical forensic case or paternity trio.

The RFLP analysis of VNTR loci was implemented for forensic casework in this country in 1988 by the Federal Bureau of Investigation (FBI) after completion of extensive validation studies *(10)*. Although VNTR identification techniques enjoyed widespread acceptance by the US legal system, the methods employed offered several serious limitations. With Southern blots being stripped and reprobed multiple times using single-locus probes, the procedure is quite labor-intensive, sometimes requiring several weeks for complete testing, and these procedures are not easily adaptable to automation—which is ultimately required of any genetic system to be used for large-scale testing and databasing. Additionally, because the VNTR-containing regions are relatively long, large amounts of nondegraded genomic DNA (50–100 ng) are required for successful analysis. This becomes especially limiting for forensic samples that are commonly old, degraded, and of limited quality.

Although some commercial laboratories continue to offer RFLP testing *(2)*, over the last decade these methods have largely given way to polymerase chain reaction (PCR)-based techniques, which have revolutionized the field of human identification. The unique ability of PCR to handle forensic samples that are of limited quality and quantity, along with the desire for rapid, high-volume processing have driven the field away from RFLP methods. The first PCR system used for forensic DNA typing was based on sequence variation at the DQα locus of the human leukocyte antigen gene cluster (HLA-DQα) *(11)*. DQα is a 242-bp region, with 4 known primary alleles and 3 additional subtypes of the 4 allele and 1 allele, making for a possible 28 detectable genotypes. Variation at this locus is detected using reverse dot blot technology whereby sequence-specific probes for each allele are immobilized on a membrane strip *(12)*. The exact type of the sample is revealed after hybridization of the amplified DNA to a specific probe on the strip, causing the enzymatic conversion of a colorless substrate (tetramethylbenzidine) to a blue precipitate. As shown in Fig. 2, the pattern of blue dots seen corresponds to the alleles present in the sample and a comparison of the pattern between two typing strips would indicate whether two samples originated from the same source or if parentage could be established. A major advantage to PCR-based methods is the potential to amplify multiple genetic regions at the same time—multiplexing. Thus, soon after the release of the original HLA-DQα kit by Cetus, Roche Molecular Systems developed and released an additional kit in which HLA-DQα typing was offered in conjuction with a second strip that was capable of detecting alleles at five additional genetic markers (LDLR, GYPA, HBGG, D7S8, and GC), all from a single amplification. The new kit, termed "AmpliType PM + DQα," was simply an expansion of the techniques used in the original HLA-DQα analysis. Although each additional locus offered less individual variation than DQα, the combined results increased the power of discrimination of the test considerably to approx 1 in 2000 *(13)*. However, this was significantly lower than the discriminatory power of a single RFLP/VNTR locus. Additionally, there were several other disadvantages to this test, including its limited ability to distinguish samples containing DNA from more than one contributor (mixture analysis). Although the AmpliType PM + DQα system is still commercially available and used for other typing applications *(14)*, it was phased out after a few years in most human identity laboratories in favor

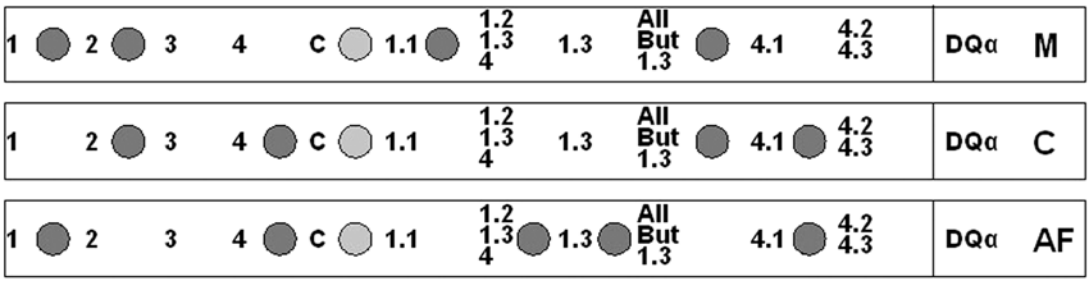


Fig. 2. DQα reverse dot blot. DNA samples from a paternity trio (M, C, AF) were amplified by PCR for the HLA-DQα locus. PCR products were hybridized to DQα strips. Each blue dot corresponds to the allele(s) shown, with the control dot (C) shown in the center and subtypes for the 1 and 4 alleles shown on the right side of the strip. This schematic of the DNA paternity results reveal the types of the M (1.1,2), C (2,4.1), and AF (1.3,4.1), which indicate that the AF is included at this locus.

of multiplexed short tandem repeat (STR) systems. Today, STRs remain the system of choice for forensic and paternity testing laboratories *(2,15)*.

Short tandem repeat loci are amendable to PCR because of their relatively small size. STR loci contain shorter repeat sequences, ranging from 2 to 7 bp in length, with a fewer number of repeats at each locus making their total length significantly shorter than VNTRs. Total length of a typical STR locus ranges between 100 and 400 bp, in comparison to RFLP regions, which can vary in length from 500 to 12,000 bp, making STRs an ideal choice for DNA that might be partially degraded. STR loci with tetranucleotide repeat sequences are most commonly chosen for use in human identification testing *(16,17)*. The analysis of these marker systems result in discrete alleles and can be compared directly to an allelic ladder, simplifying data interpretation, analysis, and comparison through computerized DNA databases. Initially, scientists employed polyacrylamide gels for STR fragment separation along with silver staining for detection of alleles *(18)*. However, this technique has been replaced by fluorescence-based detection methods, which are more sensitive, have higher throughput capacities, and have been adapted for automated detection, analysis, and storage of data *(19,20)*. Multiplex PCR of STR loci using fluorescently tagged primers allows for simultaneous detection of these fragments in a single lane (for slab-gel electrophoresis methods) or a single injection (for capillary-electrophoresis-based sequencers). Although individually each STR locus is only moderately polymorphic when compared to larger VNTR loci, multiplexing, combined with automated detection/analysis, allows for a significant decrease in time and labor, thus increasing cost savings *(15)*. Additionally, the STR loci that have been chosen for forensic and parentage testing are reasonably well distributed in most populations, increasing the power of discrimination between individuals when multiple loci are tested. More detail on the current techniques used for STR amplification, detection, and analysis will be discussed later in this chapter.

1.2. QUALITY ASSURANCE Shortly after the implementation of RFLP casework by the FBI, the Technical Working Group on DNA Analysis Methods (TWGDAM) was established to guide forensic scientists on technical development, validation, and quality assurance issues. The guidelines issued by TWGDAM served as the *only* guidelines until the DNA Advisory Board (DAB) was created and funded by the US Congress DNA Identification Act of 1994. The primary responsibilities of the DAB were to issue standards for the forensic DNA community and to implement a nationwide DNA index for law enforcement purposes similar to the database system that was already in operation by the United Kingdom's Forensic Science Service (FSS) Unit *(21)*. In 1998, the DAB issued the "Quality Assurance Standards for Forensic DNA Testing Laboratories" *(22)*, which included requirements for personnel, proficiency testing, evidence control, analytical procedures, safety, and validation. Later that same year, the FBI launched its nationwide DNA database, the Combined DNA Index System (CODIS), giving federal, state, and local crime labs the ability to exchange and compare DNA profiles electronically, linking crimes to each other and to specific convicted offenders. Shortly after the implementation of the CODIS database, the DAB issued a second somewhat overlapping set of standards specifically for laboratories processing convicted offender samples for the database, the "Quality Assurance Standards for Convicted Offender DNA Databasing Laboratories" *(23)*. Both sets of standards called for an annual audit for all forensic and/or convicted offender testing laboratories, with an *external* audit every 2 yr. As a result, the FBI joined forces with the two forensic DNA accrediting agencies—the National Forensic Science Technology Center (NFSTC) and the American Society of Crime Lab Directors/Laboratory Accreditation Board (ASCLD/LAB)—to issue a single, combined quality assurance audit document and to participate in a joint training program based on the use of the Quality Assurance Audit Document *(24)*. All laboratories wishing to submit DNA evidence to the US court system are expected to follow the aforementioned standards, including being subjected to annual auditing. Laboratories that adhere to the highest of standards will additionally request either or both ASCLD/LAB or NFSTC accreditation. See Table 1 for a list of accredited commercial laboratories offering forensic DNA casework services and/or convicted offender sample processing *(25,26)*. At the conclusion of the DAB mandate, TWGDAM converted to SWGDAM and this group is now responsible for offering technical and developmental recommendations to the forensic DNA community as well as being responsible for overseeing modifications and updates to the DAB-issued quality assurance standards.

Table 1
US Commercial Human Identification Laboratories

Laboratory	*Location*	*NFSTC*	*ASCLD/LAB*	*AABB*
American Red Cross Blood Services	*Madison, WI Portland, OR*			Δ
Analytical Genetic Testing Center, Inc.	*Denver, CO*			Δ
Blood Systems Laboratories	*Tempe, AZ*			Δ
Bode Technology Group	*Springfield, VA*	Δ	Δ	
Boston School of Medicine–Center for Human Genetics	*Boston, MA*	Δ		
BRT Laboratories	*Baltimore, MD*	Δ		
Central Indiana Regional Blood Center	*Indianapolis, IN*			Δ
Clinical Testing and Research, Inc.	*Ridgewood, NJ*			Δ
Commonwealth Biotechnologies	*Richmond, VA*	Δ		Δ
DNA Diagnostics Center	*Fairfield, OH*			Δ
DNA Diagnostic Laboratory–University of Utah	*Salt Lake City, UT*			Δ
DNA Reference Laboratory	*San Antonio, TX*	Δ		Δ
DNA Testing Laboratories	*Cincinnati, OH*			Δ
Fairfax Identity Laboratories	*Fairfax, VA*	Δ		Δ
Forensic Analytical	*Hayward, CA*	Δ		
Genelex Corporation	*Redmond, WA*			Δ
Genetica DNA Laboratories Inc.	*Cincinnati, OH*			Δ
Genetic Profiles	*San Diego, CA*		Δ	
Genetic Technologies, Inc.	*Glencoe, MO*		Δ	
GenQuest–University of Nevada	Reno, NV		Δ	
H.A. Chapman Institute of Medical Genetics	Tulsa, OK		Δ	
Identigene, Inc.	Houston, TX	Δ		Δ
Identity Genetics, Inc.	Brookings, SD	Δ		Δ
Immunogenetics/ DNA Diagnostics Laboratory–University of Alabama at Birmingham	Birmingham, AL		Δ	
Laboratory Corporation of America (LabCorp)	Burlington, NC Research Triangle, NC	Δ	Δ	
Legal Genetics, Inc.	Ocean Springs, MS		Δ	
Long Beach Genetics Esoterix	Rancho Dominguez, CA			Δ
Marshall University–Forensic Science Center	Huntington, WV	Δ		Δ
Medical College of Ohio	Toledo, OH		Δ	
Memorial Blood Centers of Minnesota	Minneapolis, MN		Δ	
Mitotyping Technologies, LLC	State College, PA	Δ		
Molecular Pathology Laboratory	Maryville, TN		Δ	
Myriad Genetics Laboratory	Salt Lake City, UT	Δ		
National Medical Services	Willow Grove, PA	Δ		
Orchid Cellmark–Germantown	Germantown, MD	Δ	Δ	
Orchid Cellmark–Nashville	Nashville, TN	Δ	Δ	Δ
Orchid GeneScreen–East Lansing	East Lansing, MI		Δ	
Orchid GeneScreen–Dallas	Dallas, TX	Δ	Δ	Δ
Orchid GeneScreen–Dayton	Dayton, OH		Δ	
Paternity Testing Corporation	Columbia, MO	Δ		Δ
ReliaGene Technologies, Inc.	New Orleans, LA	Δ	Δ	Δ
Rhode Island Blood Center	Providence, RI		Δ	
SWBIC Genetic Testing Laboratory – New Mexico State University	Las Cruces, NM		Δ	
Serological Research Institute	Richmond, CA	Δ		
University of North Texas	Ft. Worth, TX		Δ	

In 1982, the Office of Child Support Enforcement of the US Department of Health and Human Services offered grant support to the American Association of Blood Banks (AABB) to determine what tools were sufficient to determine inclusion and exclusion of parentage *(27)*. Within 3 yr, DNA technology was introduced and quickly became the mainstay technique for establishing relatedness *(28)*. Since this time, the AABB has become responsible for issuing national standards for parentage testing laboratories providing assessment and accreditation of parentage testing laboratories and collecting annual statistics regarding case volumes, trends in technologies, and mutation rates for RFLP and PCR loci (2). As of July 2003, there were

37 accredited paternity testing laboratories in the United States (Table 1) *(29)*. In addition, there are several public crime laboratories that offer paternity testing for relevant casework samples.

2. CURRENT TECHNIQUES FOR HUMAN IDENTIFICATION

2.1. EVIDENCE COLLECTION AND SAMPLE PREPARATION As with other types of clinical lab testing, the most important step for any forensic or parentage analysis is the collection of the biological sample. Successful criminal or civil cases might not be possible to salvage if evidence is mishandled during the initial steps of the investigative process. Initial steps include, but are not limited to, proper documentation (chain of custody), collection, packaging, labeling, and preservation. Individuals handling evidence samples should be properly trained on how to avoid contamination and secondary transfer of DNA samples, how to remove stains from unmovable objects, how to thoroughly dry and package each item separately, and how to use universal precautions during collections, including the use of clean gloves for each individual sample collection. Biological samples should be thoroughly air-dried, stored cold, and shipped overnight to an appropriate testing laboratory. Once DNA is extracted, it should be stored either in a refrigerator at 4°C or a freezer at –20°C. DNA can also be stored long term at –70°C. Lee et al. *(30)* provide readers with a more thorough review of DNA evidence collection guidelines covering a variety of biological materials.

The most commonly obtained reference samples for routine paternity analysis or convicted offender typing are buccal epithelial swabs, whole-blood samples, or bloodstain cards *(2)*. In contrast, the most common biological materials tested from crime scenes for forensic casework analysis are blood/bloodstains and semen/semen stains. However, a variety of sample types have successfully yielded PCR-generated DNA profiles, including bones, teeth, hairs, urine, saliva, feces, soft tissue, cigaret butts, fingerprints, and personal items *(1,15)*.

2.1.1. DNA Isolation and Quantitation Methods Prior to performing forensic or parentage analyses on biological samples, the DNA must be isolated and the DNA quantity and quality need to be evaluated to assure optimal results. In forensic analysis especially, it is desirable to complete the DNA isolation as expeditiously as possible to avoid potential degradation. Although any number of methods can be used to extract DNA from biological material, identification labs generally use either traditional organic methods or an extraction kit that has been optimized and validated for use by established labs in the forensic community. The protocols selected tend to be those that are more rapid, have fewer steps and tube transfers, those that have increased yields of high-molecular-weight DNA, and those that are adaptable to automated processing. During the extraction process, all samples to be tested for human identification purposes must be handled extremely carefully to avoid sample-to-sample contamination and the introduction of foreign DNA. For forensic casework, it is required by the DAB standards that crime scene samples be processed separate from reference samples in both time and space.

The procedure selected for DNA isolation will vary somewhat according to the type and amount of biological sample that is obtained. Chelex extraction of DNA was the first method to become very popular among forensic scientists when laboratories began switching to PCR-based typing procedures *(31)*. However, practitioners quickly discovered that other methods consistently produced cleaner DNA preparations with greater long-term stability and were more reliable for STR analysis *(32)*. Today, there are three basic methods that are commonly employed in human identification laboratories, including FTA paper processing, Qiagen QIAAmp extraction, and organic (phenol/chloroform) extraction. FTA paper was developed in the late 1980s as an absorbent cellulose-based treated paper that is stable for long-term, room-temperature DNA storage. Liquid samples containing biological material can be spotted onto FTA paper at which time, cells are lysed, and nucleic acids are immobilized within the cellulose matrix *(33)*. A small punch can be removed, washed, and added directly to PCR multiplex reactions to successfully obtain full STR profiles *(34)*. Typical sample sources collected for FTA storage/processing include whole blood, rehydrated blood stains, and buccal cells transferred from cheek swabbings *(33,35)*. More recently, FTA has become the preferred method for processing of reference samples for many laboratories because of the availability of semiautomated punching devices, such as the Wallac DBS punch and the BSD robots. Alternatively, Qiagen extractions have also been recently popularized because of their ease of use, high DNA yields *(36)*, format availability (tube or 96 wells), their optimized tissue-specific protocols and kits *(37–39)*, and their ability to be easily adapted for use on automated high-throughput liquid-handling devices. Qiagen protocols employ the use of a silica-based column over which the cellular lysate is passed. Under certain chemical conditions, DNA will adhere to the column, allowing other materials to be washed off. The final DNA sample will be eluted from the column and can be added directly to PCR reactions after proper quantitation and dilution *(32,37)*.

Although many of these methods can be modified for processing alternative samples sources, such as hairs, teeth/bone, or saliva, those sample types, which typically have very low DNA yields, are routinely processed using more traditional methods, such as phenol/chloroform (organic) extraction methods followed by Microcon filtration or alcohol precipitation *(40–42)*. Modified organic extractions are also commonly used in forensic casework for differential extraction of DNA from vaginal samples that might contain male (sperm) cells *(40,43)*. This allows for distinction between the male STR profile (sperm fraction) and female STR profile (nonsperm or epithelial fraction).

A newer method of DNA extraction that has recently been introduced to the forensic/parentage testing community, the DNA IQ System (Promega Corporation, madison, WI), could soon replace older methods that require extensive washing and are less amenable to automation. The DNA IQ System uses a specific paramagnetic resin to capture small quantities of DNA. The protocol is designed so that a maximum of approx 100 ng of DNA will be isolated, and excess DNA and cellular debris are easily removed by pipetting. After several wash steps, the DNA is eluted using heat. This system can be used with a variety of sample sources and can be easily automated *(44)*. Several

forensic laboratories have implemented this system to be used in conjunction with the Beckman Biomek 2000 liquid-handling robotic platform *(45)*.

It is especially imperative for PCR-based typing methods that DNA quantitation be determined prior to amplification, as most multiplex PCR assays require a narrow range of input template DNA (0.2–2 ng). For reference-type samples, including convicted offender and parentage samples, many labs find it sufficient to quantify DNA samples by measuring absorbance values at 260 nm or measuring fluorescence of samples after the addition of an interchelating dye (PicoGreen) using a fluorometer *(46,47)*. Both are useful for adjusting input DNA concentrations appropriately and are easily formatted in 96-well microtiter plates for automated processing. However, some extraction procedures, such as FTA paper punching and the DNA IQ System, are designed to give consistent yields on each extraction, such that individual quantitation of each sample might not be necessary for standard reference or single source samples *(33,35,44)*. Although these methods are effective for reference samples, they estimate total DNA content (microbial, human, etc.) and are thus not sufficient for samples from forensic casework.

In the Quality Assurance Standards for Forensic DNA Testing Laboratories, standard 9.3, the DAB calls for all laboratories to "follow a procedure for evaluating the quantity of the human DNA" for forensic samples to be analyzed *(22)*. The most popular method used by forensic laboratories to quantify human-specific DNA is the slot-blot procedure *(48,49)*. In this method, a small portion of each sample and several known standards are transferred to a nylon membrane followed by the addition of a human-specific probe (D17Z1). Resulting chemiluminescent or colorimetric signal intensities can then be compared to the known standards and relative measurements can be estimated *(50)*. This assay is available as a commercial product, "QuantiBlot Human DNA Quantitation Kit," from Applied Biosystems. Although this assay is steadily used by most forensic laboratories in the United States, it gives variable results from lab to lab and is labor-intensive and time-consuming, taking approx 3.5 h to perform for only 20–30 samples. Thus, forensic science researchers are currently in search of faster, more sensitive assays that could be developed for this purpose.

Several new assays for human quantitation have been described recently and are undergoing validation within the forensic science community. One approach involves the amplification of a single human-specific STR locus *(51)*. PCR products from this assay can be labeled with standard double-stranded DNA dyes (such as PicoGreen) and fluorescence measured with a fluorometer. This procedure offers the advantage that samples can be processed in a 96-well format using commonly available forensic laboratory equipment and only requires approx 1 h of analyst time. Another approach that has been frequently discussed in the literature is the quantitation of human DNA by amplification of primate-specific, multipcopy *Alu* insertional elements that are dispersed throughout the human genome. Several methods for postamplification detection have been described, including SYBR green-based real-time PCR or fluorometer detection *(52,53)*, incorporation of fluorescently labeled primers followed by fluorometer or ABI analyzer detection *(54)*, and luciferase incorporation with light detection via a luminometer ("AluQuant Human DNA Quantitation System," Promega Corporation) *(55)*. A major advantage of these methods is rapid analysis time, however, many involve the use of expensive equipment that might not be available to most forensic laboratories. At this time, it is not clear if these methods will increase accuracy and reproducibility enough to warrant their widespread implementation.

2.2. PCR AMPLIFICATION OF STR LOCI In the United States, 13 tetranucleotide STR loci have been selected for use in the national DNA indexing system (CODIS): including CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, and D21S11 (Fig. 3). These loci were selected for human identification based on several criteria *(15,56)*. All produced lower stutter rates than other STR loci and all had lower mutation rates. Additionally, all of the known alleles of the selected loci fell within the range of 90–500 bp and could be discretely and reproducibly detected when multiplexed with other markers. Loci chosen are also typically located on separate chromosomes (*see* Fig. 3) so that frequency data can be easily calculated using the product rule and linkage problems avoided. Generally, the loci selected are highly polymorphic, have observed heterozygosity levels of >70%, and have a widespread distribution of alleles among populations. Together, these things contribute to the very high discriminatory power of the core 13 STR loci. In fact, according to Chakraborty et al., the average random match probability for the combined 13 STR loci is rarer than 1 in a trillion *(57)*.

Initially, the development and characterization of the commonly used STR loci was carried out by the laboratories of Dr. Thomas Caskey (Baylor University College of Medicine) *(17,58)* and the Forensic Science Service (England) *(59,60)*. Ultimately, these STR loci were commercialized by both Applied Biosystems and the Promega Corporation—providing the forensic community with a common set of loci that could be amplified in only one or two multiplex PCR amplifications from less than 1 ng of DNA. Today, these commercially available amplification kits dictate which STRs will be tested in most forensic and parentage testing laboratories. This system allows for improved opportunities for laboratories to exchange data without the fear of inconcordance or null alleles. For a list of the commonly used commercially available human identification amplification kits, see Table 2. Kits that are currently manufactured are advantageous in that alleles can be directly detected with automated fluorescent detection systems, eliminating the need for the use of probes or postelectrophoresis staining of the gel. One primer from each set of primers included in the amplification mixture has a fluorescent label attached to it, such that alleles at overlapping loci can be detected using different wavelengths of light. The use of multiple-colored fluorescent dyes allows for the loci to be multiplexed. As shown in Fig. 4, this approach allows for the simultaneous electrophoresis and detection of 8–15 STR loci plus the Amelogenin gender locus all at once *(61)*. More information on the relevant commercially available amplification kits, STR loci, and primers can be found on the STRBase website http://www.cstl.nist.gov/biotech/strbase/ *(62)*.

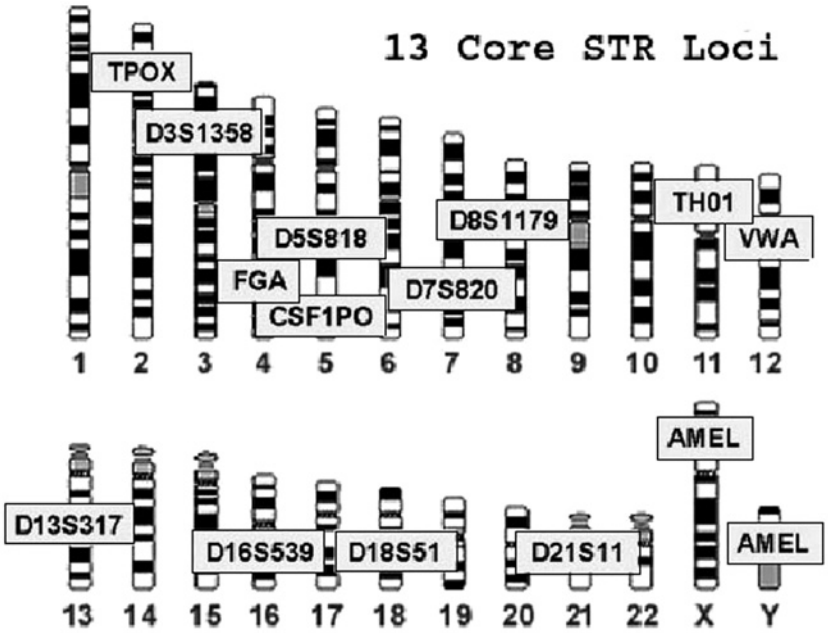


Fig. 3. Thirteen core STR loci with chromosomal positions. Thirteen core STR loci were selected by the FBI for inclusion into CODIS, the national DNA database. Shown is the approximate location of each of the autosomal STRs on their respective chromosomes and the Amelogenin gender marker on the X and Y chromosome. (Courtesy of John Butler, STRBase/National Institutes of Standards and Technologies.)

Table 2
Commercially Available Fluorescent STR Multiplexes

Name	*Vendor*	*Average power of discrimination*[a]	*STR loci included*
FFFL	Promega Corporation	1 : 2700	F13A1, FES/FPS, F13B, LPL
AmpF1STR® Blue	Applied Biosystems	1 : 5000	D3S1358, VWA, FGA
CCTv	Promega Corporation	1 : 6800	CSF1PO, TPOX, TH01, VWA
GammaSTR®	Promega Corporation	$1 : 1.7 \times 10^4$	D16S539, D7S820, D13S317, D5S818
AmpF1STR® COfiler®	Applied Biosystems	$1 : 8.4 \times 10^5$	D3S1358, D16S539, Amelogenin, TH01, TPOX CSFlPO, D7S820
PowerPlex® 1.1/PowerPlex® 1.2	Promega Corporation	$1 : 1.1 \times 10^8$	CSF1PO, TPOX, TH01, VWA, D16S539, D13S317, D7S829, D5S818
AmpF1STR® Profiler®	Applied Biosystems	$1 : 5 \times 10^9$	D3S1358, VWA, FGA, Amelogenin, TH01, TPOX, CSF1PO, D5S818, D13S317, D7S820
PowerPlex® 2.1	Promega Corporation	$1 : 8.5 \times 10^{10}$	D3S1358, TH01, D21S11, D18S51, VWA, D8S1179 TPOX, FGA, Penta E
AmpF1STR® Profiler Plus®	Applied Biosystems	$1 : 9.6 \times 10^{10}$	D3S1358, VWA, FGA, Amelogenin, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820
AmpF1STR® Profiler Plus ID®	Applied Biosystems	$1 : 9.6 \times 10^{10}$	D3S1358, VWA, FGA, Amelogenin, D8S1179, D21S11, D18S51 (+ modified primer), D5S818, D13S317, D7S820
AmpF1STR® SGM Plus®	Applied Biosystems	$1 : 5 \times 10^{12}$	D3S1358, VWA, D16S539, D2S1338, Amelogenin, D8S1179, D21S11, D18S51, D19S433, TH01, FGA
AmpF1STR® SEfiler®	Applied Biosystems	$1 : 1.3 \times 10^{13}$	D8S1179, D18S51, D21S11, FGA, TH01, VWA, Amelogenin, D2S1358, D3S1358, D16S539, D19S433, SE-33
PowerPlex® 16	Promega Corporation	$1 : 1.8 \times 10^{17}$	CSF1PO, FGA, TPOX, TH01, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, Penta D, Penta E, Amelogenin
AmpF1STR® Identifiler®	Applied Biosystems	$1 : 2.1 \times 10^{17}$	CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, vWA, Amelogenin, FGA, TH01, TPOX, D2S1338, D19S433
PowerPlex® ES	Promega Corporation	Not reported	D18S51, D21S11, TH01, D3S1358, FGA, D8S1179, VWA, Amelogenin, SE-33

[a]Approximate average reported for Caucasian population.

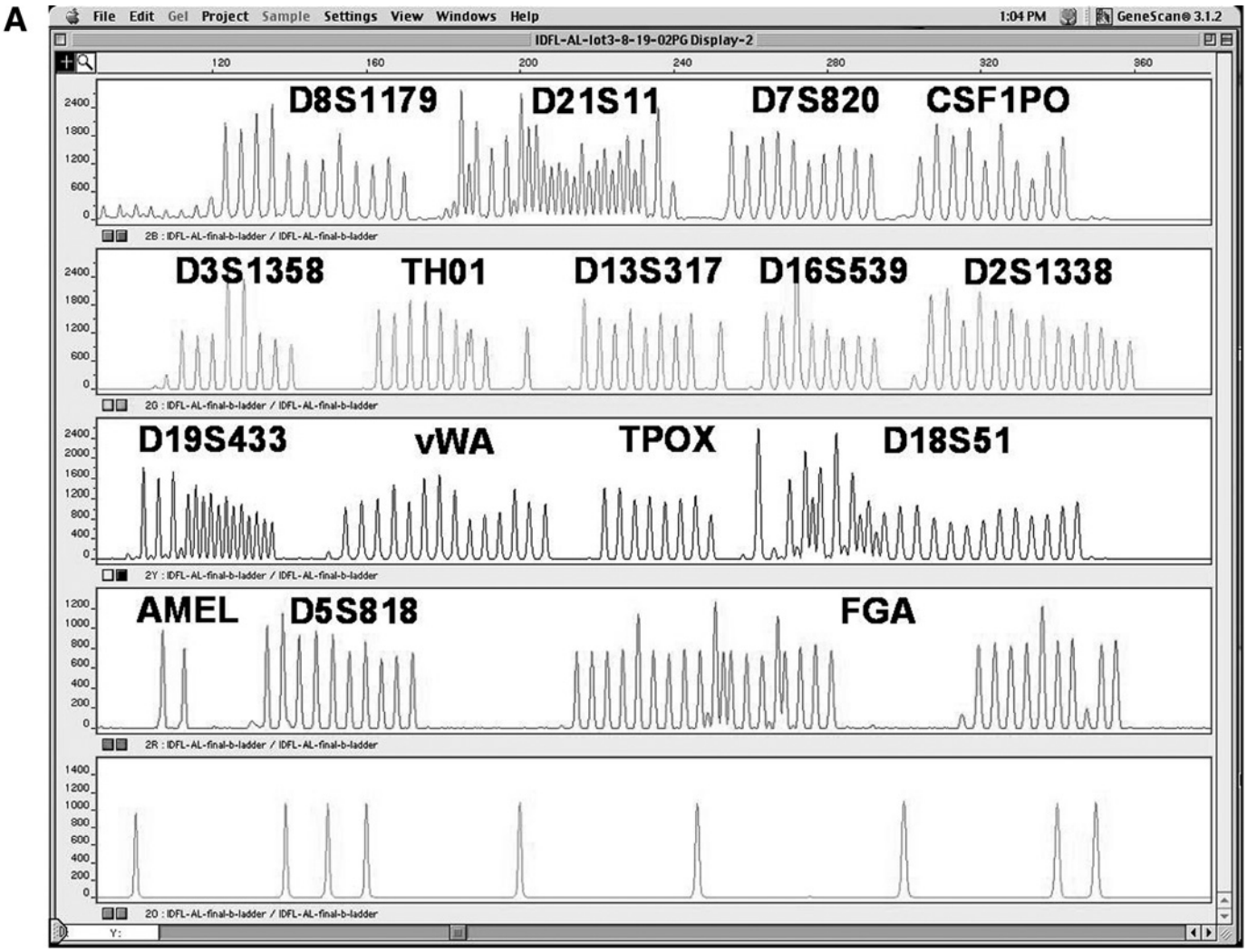


Fig. 4. AmpFlSTR® Identifiler®. **(A)** Identifiler® allelic ladder. ABI's Identifiler amplification kit provides this allelic ladder, which includes approx 200 alleles, for assignment of genotypes after PCR multiplex reactions involving 16 genetic loci and 4 fluorescent dyes. Alleles represent the 13 core STR loci along with 2 additional tetranucleotide STR loci (D2S1338 and D19S433) plus the Amelogenin gender-determining locus. The fifth dye lane (orange) shows the internal lane standard. **(B)** Identifiler DNA profile from a single individual. The top four panels show the alleles for specific loci in the corresponding dye colors. The bottom panel shows the internal lane standards in a fifth dye color (orange). (Courtesy of Applied Biosystems.)

2.3. STR FRAGMENT SEPARATION AND DATA ANALYSIS A number of different manual *and* semiautomated fluorescence detection systems exist and are used in forensic and parentage testing laboratories for STR allele determination or "profiling." However, the most popular platforms used today for STR analysis are the FMBIO II gel scanner and the ABI Prism 310/3100 Genetic Analyzers, which are fundamentally very different in their approach to fluorescent detection *(15,62–64)*. In the first system, the FMBIO II gel scanner, PCR products are run out on acrylamide gels followed by detection via rapid scanning on the FMBIO system scanner. Allele fragments are initially analyzed with the FMBIO Analysis Software, producing information on DNA fragment sizes that is stored in the associated computer. With this system, STR alleles are called using the STaR Call Genotyping Software. Although the gel pouring/loading is quite manual and labor-intensive, the minimal scanning time required for each gel allows for high-throughput processing of 400 samples per day or more *(15)*.

In contrast, ABI fluorescent detection systems are equipped with a laser beam that excites the fluorescent DNA fragments as they are electrophoresing through the capillaries, and a charge-coupled devise (CCD) camera, which captures the data for storage in the associated computer. Early models such as the ABI 373 or 377 were gel-based systems in which the samples were run through an acrylamide gel (similar to the FMBIO systems) using real-time data capture rather than postelectrophoresis data capture *(65,66)*. Unfortunately, as with other slab-gel methods, these proved to be time-consuming and labor-intensive and not as amenable to automation as other systems. For these reasons, most forensic and parentage laboratories have transitioned to capillary electrophoresis (CE)-based platforms such as the ABI Prism 310/3100 Genetic Analyzers *(67,68)*. With these models, the disadvantages of the gel-based systems are removed, as the sample bands are run through a capillary column prior to laser detection, instead of a gel. The major advantages that the CE-based platforms offer are the 96-well plate format and the

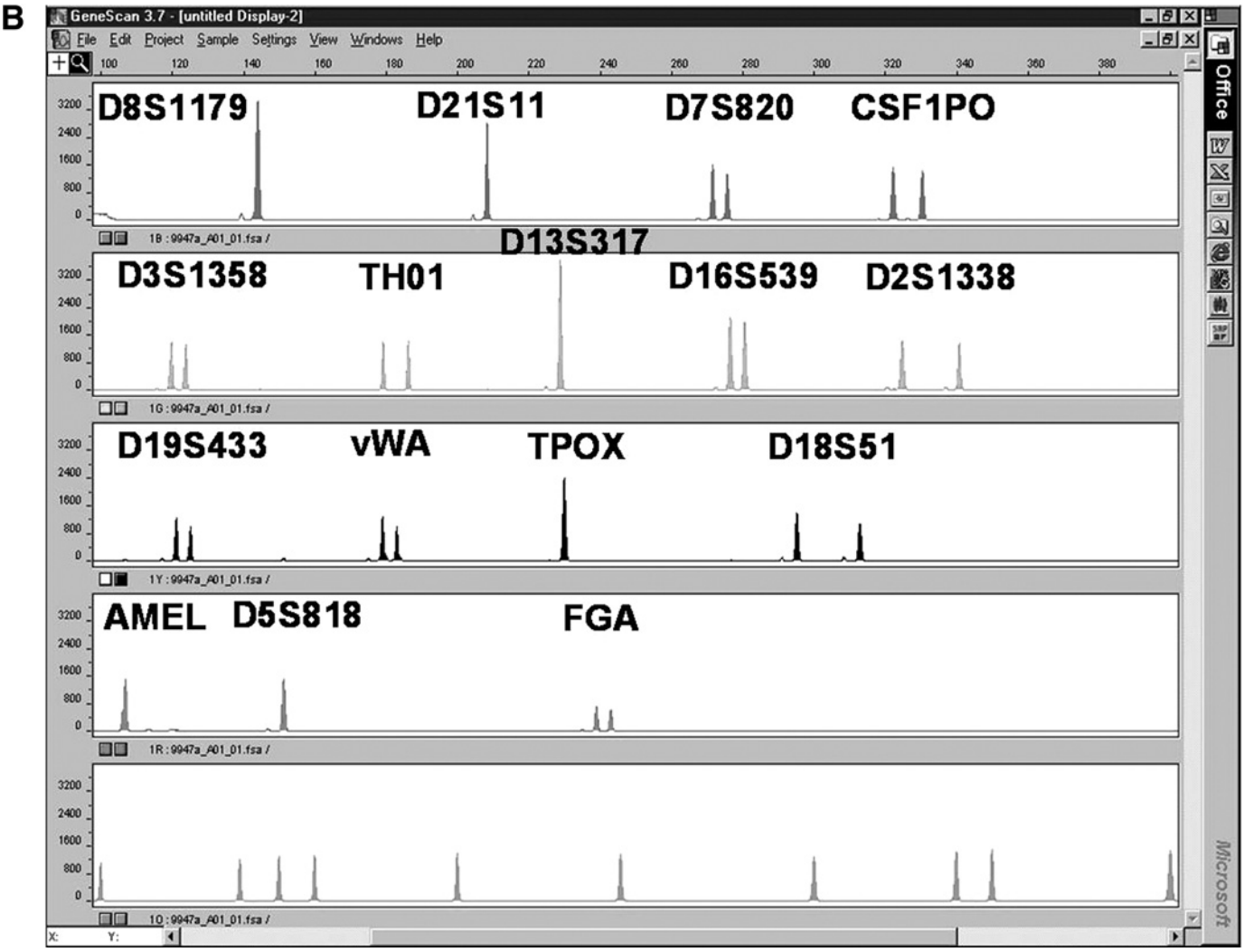


Fig. 4. (*Continued*)

capability of unattended operation. Sample plates can be prepared robotically and then simply placed on the autosampler rack. After polymer and buffer replenishment, the operator can begin the run and walk away. Data are collected via ABI Data Collection Software, fragments analyzed and sized with ABI GeneScan Software, and genotyping completed with ABI Genotyper Software. With a 310 single-capillary model, samples will be processed quite slowly at the rate of approximately one sample per 30 min *(15)*. However, throughput capability has been increased significantly with the recent releases of the ABI 3100 and 3100-Avant systems (16 and 4 capillaries, respectively). With these models, sample run time increases to approx 45 min; however, either 16 or 4 samples are being injected and analyzed at once because of their multiple-capillary array formats. With these two instruments, ABI has provided a scalable system that can meet the throughput needs of small, mid-size, or even large forensic and parentage testing labs.

Very seldom are inconsistencies reported between amplification kits and/or instrumentation technology. Several studies have completed side-by-side comparisons of the most commonly used fluorescent detection platforms (including ABI Prism 310, ABI Prism 377, and Hitachi FMBIO II) and have concluded that all of these existing technologies are sufficiently robust and precise for STR-based human identification *(64,69)*. However, these studies also show that with the validation of a new detection system, labs must independently develop interpretational guidelines to accommodate for differences in sensitivity and precision (within and between instruments) *(69)*. Concordance data for the more recently released ABI Prism 3100 and 3100-Avant models is not yet available; however, precision studies that have been completed suggest that their performance will mimic that of the other ABI Prism systems *(70)*. For the STR human ID amplification kits themselves, most have overlapping loci included to assure concordance in the data between different samples that are analyzed from the same biological source. For example, as shown in Figure 5A and B, the ABI AmpFlSTR® Profiler Plus® and COfiler® kits have the loci D3S1358, D7S820, and Amelogenin in common, whereas the Promega PowerPlex® 1.1 and PowerPlex® 2.1 have the loci TH01, TPOX, and VWA in common. However, it is important to note that this internal quality control check is lost if laboratories choose to implement the newer single-amplification 16-locus kits, such as ABI's Identifiler® or Promega's PowerPlex® 16. Most STR allele inconsistencies reported in the published

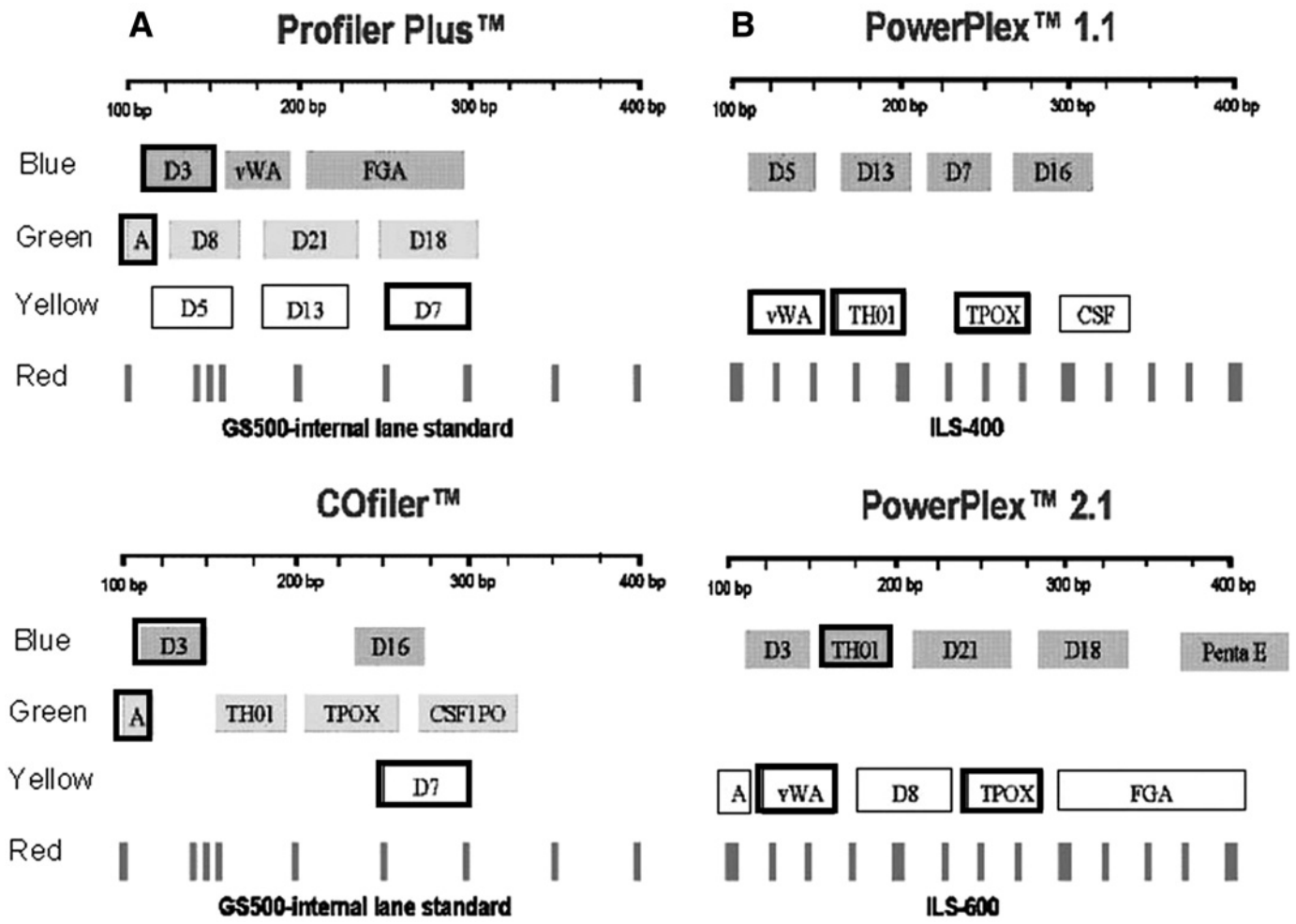


Fig. 5. Commercially available kits for core STR amplification. Most human identification laboratories are using either ABI AmpFlSTR® Profiler Plus®/COfiler® kits **(A)** or Promega PowerPlex® 1.1/PowerPlex® 2.1 kits **(B)** for amplification of the 13 core STR loci. Loci with darker boxes are present in both of the kits in that set and serve to provide confirmation for allele calls and show that samples were appropriately identified and set up. (Courtesy of John Butler, STRBase/National Institutes of Standards and Technologies.)

literature show that allele amplification differences between kits can occasionally be seen when kits produced by different manufacturers are used to test the same biological sample or when new primers are incorporated for a locus. This is usually shown to be caused by allele dropout resulting from a primer-binding site mutation specific to a certain primer *(71–73)* and has been reported to occur at frequencies as high as 0.46% of alleles typed *(73)*. This is not surprising given that each manufacturer uses their own primers that have been independently developed and tested *(15,63)*. As a result, laboratories should be careful when a single-locus, single-allele mismatch is observed between an apparent homozygote and heterozygote profile.

3. ALTERNATIVE METHODS FOR HUMAN IDENTITY TESTING

3.1. MITOCHONDRIAL DNA SEQUENCING The noncoding, so-called displacement loop ("D-loop") region of the mitochondrial genome exhibits significant sequence variation between individuals and is therefore useful for human identification purposes for certain types of single-source samples. Unlike nuclear DNA, mitochondria and their associated DNA originate in the cytoplasm of the egg that formed the zygote and, thus, mitochondrial DNA (mtDNA) is maternally inherited, representing only the female ancestry of an individual. mtDNA can be advantageous because mitochondria are present in nonnuclear tissues (such as hair shafts) and are present in high copy numbers within each cell. Therefore, mtDNA can often be obtained in significant quantities even when nuclear DNA is absent, highly degraded, or too low for successful STR analysis *(74)*. For these same reasons, mtDNA sample extractions must be performed in a very clean, isolated laboratory environment using extreme caution to avoid contamination *(15)*.

Two hypervariable regions within the D-loop—HV1 and HV2—are of particular interest, as they are the regions that contain the majority of the polymorphic sites *(75)*. Approximately 610 bp are routinely evaluated and compared to the Anderson reference sequence *(76)* via traditional cycle-sequencing methods *(77)* most often utilizing the same gel and/or CE-based technologies that have been discussed earlier in this chapter *(78,79)*. However, quicker methods for rapidly screening mtDNA variation are being developed that might be useful for laboratories that find traditional base sequencing to be too expensive and cumbersome. These include minisequencing *(80,81)*, restriction enzyme digestion of HV1 amplicons with CE separation *(82)*,

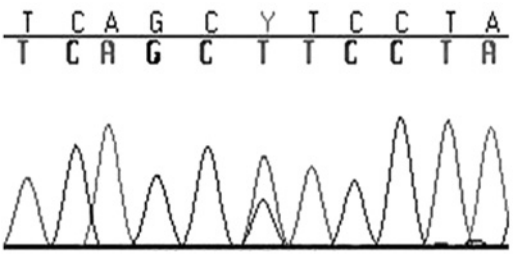


Fig. 6. mtDNA sequence heteroplasmy. Sequencing results spanning 11 nucleotide positions from the HVI (control) region of mtDNA are shown here. The sample shown has both a C and a T at the middle position indicating heteroplasmy (designated within the sequence as Y). (Courtesy of Tom Parsons and Jennifer O'Callaghan, Armed Forces DNA Identification Laboratory.)

denaturing high-performance liquid chromatography *(83)*, and immobilized sequence-specific oligonucleotide probes for HVI/HVII polymorphic sites *(84)*. Regardless of the typing methods employed, mtDNA data interpretation and sequence comparisons can be complicated by the presence of heteroplasmy—the presence of more than one base at a site in the mtDNA sequence of an individual (Fig. 6). However, recent studies have shown that heteroplasmy can actually improve the discriminating power of the analysis and strengthen the power of the mtDNA evidence *(85,86)*.

Mitochondrial DNA analysis will continue to play an important role in forensic and missing persons investigations, particularly those that involve hair samples, skeletal remains, or other sample types that are highly degraded and fail to yield successful nuclear STR profiles. Casework backlogs for mtDNA can be expected to increase because of the recent addition of a mtDNA sequence index within the national DNA database (CODIS). In addition, interpretational guidelines have now been issued by SWGDAM *(87)* in order to provide a framework for laboratories that wish to validate mtDNA sequencing and offer this as an alternative testing procedure. Currently, there are several labs that offer mtDNA sequencing, including Mitotyping Technologies *(88)* and Bode Technologies (Springfield, VA), as well as most federal laboratories (FBI, AFDIL, etc.).

3.2. Y-CHROMOSOME STR ANALYSIS The use of the Y chromosome has long been a popular method for tracing human evolution through male lineages and, more recently, for the establishment of paternity when a male offspring is in question *(89)*. In fact, Y-chromosome markers were used in 1998 to link modern-day descendants of Thomas Jefferson and the youngest son of his slave Sally Hemmings *(90)*. However, until recently, the use of the Y chromosome in *forensic* investigations has been limited because of the lack of polymorphic sites and its very low discriminatory power. The human Y chromosome is present in normal males as a single copy that is paternally inherited without any recombination. Thus, its genetic markers are not as variable as other nuclear DNA markers and many are needed to obtain a sufficiently high discriminatory power. However, as shown in Fig. 7, since the mid-1990s, a number of Y-chromosome STRs have been described *(91–93)* and they are now being developed into multiplex assays *(94–104)*, some of which can detect up to 20 Y STR loci at once. Thus far, results demonstrate that Y-STR haplotyping in multiplexed formats is sufficiently reproducible and robust for forensic casework applications *(105,106)*.

Y-STRs are particularly useful in forensic applications *(107)* because the presence of female cells, which are often found in sexual assault case samples, does not interfere with the analysis of the male profile *(108)*. Typically, differential extraction procedures are carried out on sexual assault samples to differentially lyse and separate the sperm (male) and nonsperm cells. However, identification via autosomal STR analysis might not always be successful even after this separation *(109)*. Y-STRs might be a useful alternative in these cases. In fact, the more laborious differential extraction procedures are not necessary at all for Y-STR amplification and analysis even when only low levels of male DNA are present in a high background of female DNA. In addition to this advantage, Y-STR analysis might be more attractive to forensic laboratories because the separation technologies and analysis procedures are similar, if not identical, to those used for autosomal STR processing. It is expected that the use of Y-STR markers for sexual assault and occasional paternity cases will continually increase as nomenclature and reporting criteria are standardized *(110)* and additional multiplexes are commercialized and made available to all forensic laboratories.

4. FUTURE DIRECTIONS

Despite the power that molecular biology and DNA have brought to the forensic sciences, there are limitations that warrant the continuation of research and development in search of new marker systems and more accurate, more efficient techniques. In fact, the US Congress, through the National Institutes of Justice (NIJ) Investigative and Forensic Sciences Division, has continued to increase funding annually for critical research and development in the area of forensic DNA. Specifically, NIJ's funding objectives for FY 2003 included the development of new marker systems for revelation of additional discriminatory information, exploration of nonhuman DNA markers, and development of tools and technologies that can offer faster, more reliable, less costly analysis and interpretation of DNA evidence, including those tools that can increase the success rate of aged, degraded, or otherwise limited biological samples *(111)*.

Short tandem repeats are undoubtedly the most important genetic markers in forensic science and human identification largely because they form the basis of national DNA databases throughout the world *(21)*. However, with the addition of mitochondrial sequencing indices to the US DNA database (CODIS) and the newly available Y-STR multiplex kits, one can only speculate that the demand for testing in these areas will increase in the very near future. In addition, marker systems are currently being explored that could be used to inform investigators of some of the physical characteristics of the perpetrator, such as race and hair or eye color *(112–114)*. In fact, recently a DNA test offered by DNAPrint Genomics (Sarasota, FL) showing African-American ancestry assisted authorities in tracking down a serial killer in Louisiana *(115)*. Nonhuman DNA work has also proven to be helpful in a number of recent investigations *(116,117)*. Domestic animals cohabitate with humans and can deposit hair that can be used to link a suspect

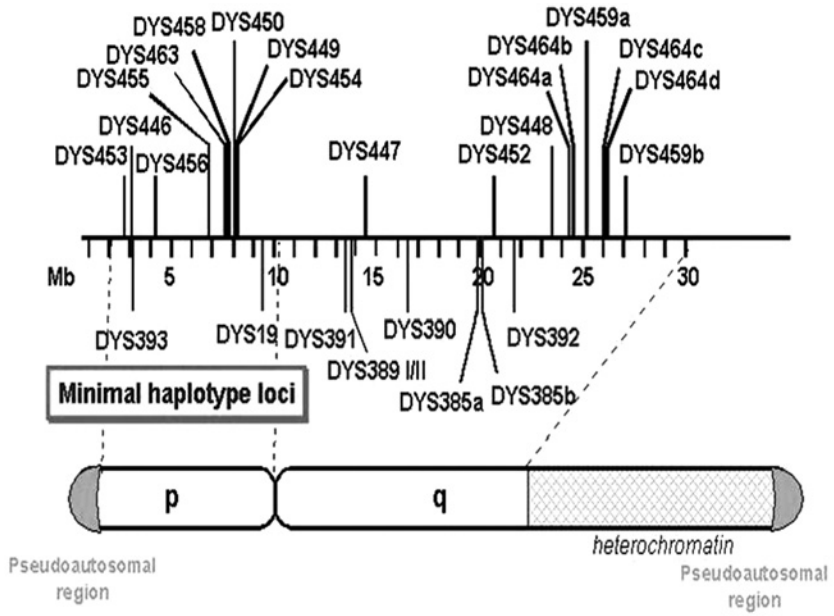


Fig. 7. Y-Chromosome STR markers. A number of Y-STRs have been described and might be potential loci for identification or lineage purposes. Several are shown on this schematic along with their relative location on the Y chromosome. The minimal haplotype loci, those required to provide minimal discrimatory power for identification, are shown on the bottom of the figure. (Courtesy of John Butler, STRBase/National Institutes of Standards and Technologies.)

to a crime scene. Both feline and canine STRs have been described and characterized *(118,119)* and, recently, an STR multiplex, "MeowPlex," as well as a feline-specific PCR quantitation assay have been developed for identification of domestic cats *(120,121)*. Savolainen et al. *(122)* have even described repeat-type sequence variation that is composed of three 10-bp repeat regions in dog mtDNA that could help to increase exclusionary power in investigations that involve domestic dogs or wolves.

Of course, the major disadvantage in other novel marker systems for personal identification outside of the core STR loci is that they cannot be used in conjunction with any of the federal, state, or local DNA databases. In the United States alone, there are now over 1.5 million profiles in the CODIS system, based entirely on the 13 core STR markers *(123)*. Thus, there would be tremendous implications if it was decided to move to a new genetic marker system. Significant advantages in terms of cost, reliability, and ease of use would have to be demonstrated. Currently, the only alternative genetic system contender is single-nucleotide polymorphism (SNP) analysis. Recently, there have been undoubted advances in SNP-based typing, which is fast becoming a powerful technology in other areas of medical genetics. SNPs are more abundant in the human genome than STRs and thousands have been characterized in linkage studies that attempt to track a variety of genetic diseases *(124,125)*. They are present on all autosomes as well as the Y chromosome *(126)*. In the forensic community, SNPs might be attractive for several reasons. First, SNPs can be multiplexed more easily and to a higher level than STRs. Also, size separation is not needed, so automated sample processing and data analysis are more amenable to automation. For example, a number of microarray platforms have been developed that can analyze thousands of SNPs simultaneously *(127,128)*. Finally, SNP PCR products can be less than 100 bp in size, which make them more efficient for processing highly degraded DNA samples than STRs, which can have amplicons as large as 400 bp *(129)*. Although these advantages are persuasive to some, there continue to be arguments against the transition to SNP typing in the forensic community based largely on the expense of changing the current standardized STR database systems worldwide and the inability of SNPs to differentiate mixed samples, particularly those consisting of more than two individuals. In addition, at least 50 SNPs will need to be analyzed to allow sufficient discriminatory purposes that will likely require a change in analysis platforms to higher-throughput technologies throughout all forensic laboratories. Although this type of change might be welcomed by some larger laboratories, most forensic DNA laboratories in this country are small and would not be able to immediately accommodate this type of technology transfer. However, SNP occurrence is known to vary widely among different human populations *(130)*, leaving open the possibility of using SNPs as markers of phenotypic interest for the development of investigative leads.

Whereas it is unlikely that STRs will be replaced soon with other genetic systems for personal identification, it is reasonable to expect changes in the platforms used for STR analysis and concurrent developments in instrument miniaturization and portability that might allow for STR analysis at the crime scene. Continued advancements in automated liquid-handling devices are also needed for faster, more efficient DNA extraction, quantitation, PCR setup, PCR, and post-PCR setup methods. Subsequently, the validation and acceptance of expert data interpretation systems such as TrueAllele (Cybergenetics) and SureLock ID (Myriad) will be critical for processing the steadily increasing backlogs that exist in forensic DNA labs. It has been reported that incorporating these types of programs

into routine STR analysis can possibly result in a 40% or more time savings *(15)*.

Researchers will also continue to focus on development of faster, less expensive, more robust methods that could be used to successfully analyze a wider range of biological samples. One exciting area of research that is ongoing in many laboratories is modification of procedures to allow for successful analysis of DNA profiles from very small quantities of DNA that might be highly degraded. Commercially available multiplexes are often designed to work at their optimum efficiency when approx 1 ng of DNA is analyzed with 28–32 cycles of amplification. However, samples that are extremely limited or samples that withstand harsh crime scene climates or long-term improper storage might not provide enough high-quality DNA for this type of analysis. Several articles have described the application of low-copy-number (LCN) DNA profiling for these types of sample, whereby PCR cycles are increased often in combination with nested PCR strategies to improve the sensitivity of standard multiplex STR amplifications *(131–135)*. However, several negative consequences of this method have been described, such as increased appearance of PCR artifacts, preferential amplification of certain alleles, and sporadic contamination from unassociated DNA that might be present on the crime scene ("allele drop-in") *(136)*. Thus, LCN analysis often requires complex data interpretations that might not be realistic for many laboratories. A more promising approach with a similar goal might be whole-genome amplification (WGA). WGA has been widely used to replicate large regions of the human genome in preimplantation genetics and diagnosis for embryonic disease, often performed from a single cell that is biopsied 2–3 d after fertilization. Several methods for WGA have been described for precursor amplification prior to genetic testing *(137)*. These include degenerate oligonucleotide primed (DOP) PCR *(138,139)*, primer extension preamplification *(140)*, and multiple displacement amplification *(141)*. Although any of these methods could be modified for use prior to multiplexed STR or SNP-based genotyping, DOP PCR might be the most promising for challenging forensic samples. In 1996, Cheung and Nelson described the use of DOP PCR for amplification of less than 1 ng of DNA that was successfully used for multiplex typing of several tetranucleotide STRs *(142)*.

Molecular biology has revolutionized the fields of human identification and forensic science. Techniques that have been described herein have led to massive changes in how parental lineage is proven, historical investigations are pursued, and in the way that forensic science is used by the police. Successful DNA casework and databasing has led to an expansion in the types of case that pursue biological evidence analysis to include most serious crimes or all felons (such as burglary and arson), some petty crimes (such as drug possession), and even arrestee testing (for select offenses). By increasing detection rates and acting as a deterrent, DNA analysis will continue to actively solve criminal cases and potentially decrease crime rates. For these reasons, molecular biologists must continue to contribute as objective participants in the legal process and continue vigorously pursuing research for more ideal techniques and practices.

ACKNOWLEDGMENTS

The author would like to thank Andrew E. Carson, Special Agent with US Army Criminal Investigative Command for providing the AFDIL information on the September 11th attacks. In addition, the author wishes to thank Susan Greenspoon of the Virginia Division of Forensic Sciences for critical review of this manuscript.

REFERENCES

1. Carson, A. E., Boyer, D., and Smith, B. Identification of the heroes of United Flight 93. Armed Forces DNA Identification Laboratory, Armed Forces Institute of Pathology. Personal communication, 2003.
2. Parentage Testing Program Unit. Annual Report Summary for Testing in 2001. American Association of Blood Banks, Bethesda, MD, 2002.
3. Shutler, G. G., Gagnon, P., Verret, G., et al. Removal of a PCR inhibitor and resolution of DNA STR types in mixed human-canine stains from a five year old case. *J. Forensic Sci.* **44:**623–626, 1999.
4. Branicki, W., Kupiec, T., and Pawlowski, R. Validation of cytochrome b sequence analysis as a method of species identification. *J. Forensic Sci.* **48:**83–87, 2003.
5. Miller Coyle, H., Ladd, C., Palmbach, T., and Lee, H. C. The Green Revolution: botanical contributions to forensics and drug enforcement. *Croat. Med. J.* **42:**340–345, 2001.
6. Wells, J. D., Introna, F., Jr., Di Vella, G., Campobasso, C. P., Hayes, J., and Sperling, F. A. H. Human and insect mitochondrial DNA analysis from maggots. *J. Forensic Sci.* **46:**685–687, 2001.
7. Jeffreys, A. J., Wilson, V., and Thein, S. L. Hypervariable 'minisatellite' regions in human DNA. *Nature* **314:**67–73, 1985.
8. Jeffreys, A. J., Wilson, V., and Thein, S. L. Individual-specific 'fingerprints' of human DNA. *Nature* **316:**76–79, 1985.
9. Klevan, L., Horton, L., Carlson, D. P., and Eisenberg, A. J. Chemiluminescent detection of DNA probes in forensic analysis. *Electrophoresis* **16:**1553–1558, 1995.
10. Budowle, B., Baechtel, F. S., Comey, C. T., Giusti, A. M., and Klevan, L. Simple protocols for typing forensic biological evidence: chemiluminescent detection for human DNA quantitation and restriction fragment length polymorphism (RFLP) analyses and manual typing of polymerase chain reaction (PCR) amplified polymorphisms. *Electrophoresis* **16:**1559–1567, 1995.
11. Blake, E., Mihalovich, J., Higuchi, R., Walsh, P. S., and Erlich, H. Polymerase chain reaction (PCR) amplification and human leukocyte antigen (HLA)-DQ alpha oligonucleotide typing on biological evidence samples: casework experiences. *J. Forensic Sci.* **37:**700–726, 1992.
12. Kawasaki, E., Saiki, R., and Erlich, H. Genetic analysis using polymerase chain reaction-amplified DNA and immobilized oligonucleotide probes: reverse dot-blot typing. *Methods Enzymol.* **218:**369–381, 1993.
13. Rudin, N. and Inman, K. *An Introduction to Forensic DNA Analysis*. CRC Press, Boca Raton, FL, 2001.
14. Giroti, R. I., Biswas, R., and Mukherjee, K. Restriction fragment length polymorphism and polymerase chain reaction: HLA-DQA1 and polymarker analysis of blood samples from transfusion recipients. *Am. J. Clin. Pathol.* **118:**382–387, 2002.
15. Butler, J. M. *Forensic DNA Typing: Biology and Technology Behind STR Markers*. Academic, San Diego, CA, 2001.
16. Kimpton, C., Fisher, D., Watson, S., et al. Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci. *Int. J. Legal Med.* **106:**302–311, 1994.
17. Hammond, H. A., Jin, L., Zhong, Y., Caskey, C. T., and Chakraborty, R. Evaluation of 13 short tandem repeat loci for use in personal identification applications. *Am. J. Hum. Genet.* **55:**175–189, 1994.
18. Kloosterman, A. D., Budowle, B., and Daselaar, P. PCR-amplification and detection of the human D1S80 VNTR locus. Amplification conditions, population genetics and application in forensic analysis. *Int. J. Legal Med.* **105:**257–264, 1993.
19. Fregeau, C. J. and Fourney, R. M. DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. *Biotechniques* **15:**100–119, 1993.

20. Gill, P., Kimpton, C. P., Urquhart, A., et al. Automated short tandem repeat (STR) analysis in forensic casework—a strategy for the future. *Electrophoresis* **16:**1543–1552, 1995.
21. Gill, P. Role of short tandem repeat DNA in forensic casework in the UK—past, present, and future perspectives. *Biotechniques* **32:**366–385, 2002.
22. DNA Advisory Board. Quality Assurance Standards for Forensic DNA Testing Laboratories. *Forensic Sci. Commun.* [online] **2,** 2000.
23. DNA Advisory Board Quality Assurance Standards for Convicted Offender DNA Databasing Laboratories. *Forensic Sci. Commun.* [online] **2,** 2000.
24. FBI Director. Quality Assurance Audit for Forensic DNA and Convicted Offender DNA Databasing Laboratories. *Forensic Sci. Commun.* [online] **3,** 2001.
25. Laboratory Accreditation Board. *Accredited Forensic Science Laboratories*. American Society of Crime Laboratory Directors, Garner, NC, 2003.
26. DNA Accreditation NFSTC. *Directory of Accredited Private DNA Laboratories*. National Forensic Science Technology Center, Largo FL, 2003.
27. Polesky, H. F. Impact of molecular (DNA) testing on determination of parentage. *Arch. Pathol. Lab Med.* **123:**1060–1062, 1999.
28. Baird, M. L. Use of DNA identification for forensic and paternity analysis. *J. Clin. Lab Anal.* **10:**350–358, 1996.
29. Parentage Testing Accreditation Program. *AABB Accredited Parentage Testing Facilities*. American Association of Blood Banks, Bethesda, MD, 2003.
30. Lee, H. C., Ladd, C., Scherczinger, C. A., and Bourke, M. T. Forensic applications of DNA typing: part 2: collection and preservation of DNA evidence. *Am. J. Forensic Med. Patholol.* **19:**10–18, 1998.
31. Walsh, P. S., Metzger, D. A., and Higuchi, R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* **10:**506–513, 1991.
32. Greenspoon, S. A., Scarpetta, M. A., Drayton, M. L., and Turek, S. A. QIAamp spin columns as a method of DNA isolation for forensic casework. *J. Forensic Sci.* **43:**1024–1030, 1998.
33. Burgoyne, L. A. Solid medium and method for DNA storage. US Patent 5,496,562, 1996.
34. Krenke, B. E., Tereba, A., Anderson, S. J., et al. Validation of a 16-locus fluorescent multiplex system. *J. Forensic Sci.* **47:**773–785, 2002.
35. Burgoyne, L. A. Convenient DNA collection and processing: disposable toothbrushes and FTA paper as a non-threatening buccal-cell collection kit compatible with automatable DNA processing, in Proceedings of the Eighth International Symposium on Human Identification, Madison, WI, p. 153, 1997.
36. Scherczinger, C. A., Bourke, M. T., Ladd, C., and Lee, H. C. DNA extraction from liquid blood using QIAamp. *J. Forensic Sci.* **42:**893–896, 1997.
37. Qiagen. *QIAamp DNA Mini Kit and QIAamp DNA Blood Mini Kit Handbook*. Qiagen. Valencia, CA, 2003.
38. Vandenberg, N. and van Oorschot, R. A. Extraction of human nuclear DNA from feces samples using the QIAamp DNA stool mini kit. *J. Forensic Sci.* **47:**993–995, 2002.
39. Sinclair, K. and McKechnie, V. M. DNA extraction from stamps and envelope flaps using QIAamp and QIAshredder. *J. Forensic Sci.* **45:**229–230, 2000.
40. Budowle, B., Smith, J. A. L., Moretti, T., and DiZinno, J. *DNA Typing Protocols: Molecular Biology and Forensic Analysis*. Biotechniques, Westborough, MA, 2000.
41. Budimlija, Z. M., Prinz, M. K., Zelson-Mundorff, A., et al. World Trade Center human identification project: experiences with individual body identification cases. *Croat. Med. J.* **44:**259–263, 2003.
42. Walsh, D. J., Corey, A. C., Cotton, R. W., et al. Isolation of deoxyribonucleic acid (DNA) from saliva and forensic sciences samples containing saliva. *J. Forensic Sci.* **37:**387–395, 1992.
43. Yoshida, K., Sekiguchi, K., Mizuno, N., et al. The modified method of two-step differential extraction of sperm and vaginal epithelial cell DNA from vaginal fluid mixed with semen. *Forensic Sci. Int.* **72:**25–33, 1995.
44. Promega. DNA IQ System—Small Sample Casework Protocol. Promega Corporation, Madison, WI, 2002.
45. Greenspoon, S. and Ban, J. Robotic extraction of mock sexual assault samples using the Biomek 2000 and the DNA IQ system. *Profiles DNA* **5:**3–5, 2002.
46. Hopwood, A., Oldroyd, N., Fellows, S., Ward, R., Owen, S. A., and Sullivan, K. Rapid quantification of DNA samples extracted from buccal scrapes prior to DNA profiling. *Biotechniques* **23:**18–20, 1997.
47. Hopwood, A. J., Mannucci, A., and Sullivan, K. M. DNA typing from human faeces. *Int. J. Legal Med.* **108:**237–243, 1996.
48. Waye, J. S., Presley, L. A., Budowle, B., Shutler, G. G., and Fourney, R. M. A simple and sensitive method for quantifying human genomic DNA in forensic specimen extracts. *Biotechniques* **7:**852–855, 1989.
49. Waye, J. S., Michaud, D., Bowen, J. H., and Fourney, R. M. Sensitive and specific quantification of human genomic deoxyribonucleic acid (DNA) in forensic science specimens: casework examples. *J. Forensic Sci.* **36:**1198–1203, 1991.
50. Walsh, P. S., Varlaro, J. and Reynolds, R. A rapid chemiluminescent method for quantitation of human DNA. *Nucleic Acids Res.* **20:**5061–5065, 1992.
51. Fox, J. C., Cave, C. A., and Schumm, J. W. Development, characterization, and validation of a sensitive primate-specific quantification assay for forensic analysis. *Biotechniques* **34:**314–322, 2003.
52. Walker, J. A., Gail, K. E., Xing, J., Shewale, J., Sinha, S. K., and Batzer M. A. Human DNA quantitation using *Alu* element-based polymerase chain reaction. *Anal. Biochem.* **315:**122–128, 2003.
53. Nicklas, J. A. and Buel, E. Development of an Alu-based, QSY 7-labeled primer PCR method for quantitation of human DNA in forensic scamples. *J. Forensic Sci.* **48:**282–291, 2003.
54. Sifis, M. E., Both, K., and Burgoyne, L. A. A more sensitive method for the quantitation of genomic DNA by *Alu* amplification. *J. Forensic Sci.* **47:**589–592, 2002.
55. Mandrekar, M. N., Erickson, A. M., Kopp, K., et al. Development of a human DNA quantitation system. *Croat. Med. J.* **42:**336–339, 2001.
56. Gill, P., Urquhart, A., Millican, E., et al. A new method of STR interpretation using inferential logic- development of a criminal intelligence database. *Int. J. Legal Med.* **109:**14–22, 1996.
57. Chakraborty, R., Stivers, D.N., Su, B., Zhong, Y., and Budowle, B. The utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems. *Electrophoresis* **20:**1682–1696, 1999.
58. Edwards, A., Civitello, A., Hammond, H. A., and Caskey, C. T. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am. J. Hum. Genet.* **49:**746–756, 1991.
59. Kimpton, C. P., Gill, P., Walton, A., Urquhart, A., Millican, E. S., and Adams, M. Automated DNA profiling employing multiplex amplification of short tandem repeat loci. *PCR Methods Appl.* **3:**13–22, 1993.
60. Urquhart, A., Kimpton, C. P., Downes, T. J., and Gill, P. Variation in short tandem repeat sequences—a survey of twelve microsatellite loci for use as forensic identification markers. *Int. J. Legal Med.* **107:**13–20, 1994.
61. Buel, E., Schwartz, M. B., and LaFountain, M. J. Capillary electrophoresis STR analysis: comparison to gel-based systems. *J. Forensic Sci.* **43:**164–170, 1998.
62. Ruitberg, C. M., Reeder, D. J., and Butler, J. M. STRBase: a short tandem repeat DNA database for the human identity testing community. *Nucleic Acids Res.* **29:**320–322, 2001.
63. Carey, L. and Mitnik, L. Trends in DNA forensic analysis. *Electrophoresis* **23:**1386–1397, 2002.
64. Moretti, T. R., Baumstark, A. L., Defenbaugh, D. A., Keys, K. M., Smerick J. B., and Budowle, B. Validation of short tandem repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. *J. Forensic Sci.* **46:**647–660, 2001.
65. Frazier, R. R., Millican, E. S., Watson, S. K., et al. Validation of the Applied Biosystems Prism 377 automated sequencer for the forensic short tandem repeat analysis. *Electrophoresis* **17:**1550–1552, 1996.
66. Fregeau, C. J., Bowen, K. L., and Fourney, R. M. Validation of highly polymorphic fluorescent multiplex short tandem repeat

systems using two generations of DNA sequencer. *J. Forensic Sci.* **44:**133–166, 1999.
67. Lazaruk, K., Walsh, P. S., Oaks, F., et al. Genotyping of forensic short tandem repeat (STR) systems based on sizing precision in a capillary electrophoresis instrument. *Electrophoresis* **19:**86–93, 1998.
68. Moretti, T. R., Baumstark, A. L., Defenbaugh, D. A., Keys, K. M, Brown, A. L., and Budowle, B. Validation of STR typing by capillary electrophoresis. *J. Forensic Sci.* **46:**661–676, 2001.
69. Tomsey, C. S., Kurtz, M., Kist, F., Hockensmith, M., and Call, P. Comparison of PowerPlex 16, PowerPlex 1.1/2.1, and ABI AmpflSTR Profiler Plus/COfiler for forensic use. *Croat. Med. J.* **42:**239–243, 2001.
70. Sgueglia, J. B., Geiger, S., and Davis, J. Precision studies using the ABI Prism 3100 genetic analyzer for forensic DNA analysis. *Anal. Bioanal. Chem.* [online], **376:**1247–1254, 2003.
71. Budowle, B., Masibay, A., Anderson, S. J., et al. STR primer concordance study. *Forensic Sci. Int.* **124:**47–54, 2001.
72. Budowle, B. and Sprecher, C. J. Concordance study on population database samples using the PowerPlex 16 kit and AmpFlSTR Profiler Plus kit and AmpFlSTR COfiler kit. *J. Forensic Sci.* **46:**637–641, 2001.
73. Alves, C., Amorim, A., Gusmao, L., and Pereira L. VWA STR genotyping: further inconsistencies between Perkin-Elmer and Promega kits. *Int. J. Legal Med.* **115:**97–99, 2001.
74. Holland, M. M. and Parsons, T. J. Mitochondrial DNA sequence analysis—validation and use for forensic casework. *Forensic Sci. Rev.* **11:**22–50, 1999.
75. Butler, J. M. and Levin, B. C. Forensic applications of mitochondrial DNA. *Trends Biotechnol.* **16:**158–162, 1998.
76. Anderson, S., Bankier A. T., Barrell B. G., et al. Sequence and organization of the human mitochondrial genome. *Nature* **290:**457–465, 1981.
77. Carracedo, A., D'Aloja, E., Dupuy, B., et al. Reproducibility of mtDNA analysis between laboratories: a report of the European DNA Profiling Group (EDNAP). *Forensic Sci. Int.* **97:**165–170, 1998.
78. Wilson, M. R., DiZinno, J. A., Polanskey, D., Replogle, J., and Budowle, B. Validation of mitochondrial DNA sequencing for forensic casework analysis. *Int. J. Legal Med.* **108:**68–74, 1985.
79. Stewart, J. E., Aagaard, P. J., Pokorak, E. G., Polanskey, D., and Budowle, B. Evaluation of a multicapillary electrophoresis instrument for mitochondrial DNA typing. *J. Forensic Sci.* **48:**571–580, 2003.
80. Tully, G., Sullivan, K. M., Nixon, P., Stones, R. E., and Gill, P. Rapid detection of mitochondrial sequence polymorphisms using multiplex solid-phase fluorescent minisequencing. *Genomics* **34:**107–113, 1996.
81. Gabriel, M. N., Huffine, E. F., Ryan, J. H., Holland, M. M., and Parsons, T. J. Improved mtDNA sequence analysis of forensic remains using a "mini-primer set" amplification strategy. *J. Forensic Sci.* **46:**247–253, 2001.
82. Butler, J. M., Wilson, M. R., and Reeder, D. J. Rapid mitochondrial DNA typing using restriction enzyme digestion of polymerase chain reaction amplicons followed by capillary electrophoresis separation with laser-induced fluorescence detection. *Electrophoresis* **19:**119–124, 1998.
83. LaBerge, G. S., Shelton, R. J., and Danielson, P. B. Forensic utility of DNA mitochondrial DNA analysis based on denaturing high-performance liquid chromatography. *Croat. Med. J.* **44:**281–288, 2003.
84. Gabriel, M. N., Calloway, C. D., Reynolds, R. L., and Primorac, D. Identification of human remains by immobilized sequence-specific oligonucleotide probe analysis of mtDNA hypervariable regions I and II. *Croat. Med. J.* **44:**293–298, 2003.
85. Ivanov, P. L., Wadhams, M. J., Roby, R. K., Holland, M. M., Weedn, V. W., and Parsons, T. J. Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Ramanov establishes the authenticity of the remains of the Tsar Nicholas II. *Nat. Genet.* **12:**417–420, 1996.
86. Grzybowski, T. Extremely high levels of human mitochondrial DNA heteroplasmy in single hair roots. *Electrophoresis* **21:**548–553, 2000.
87. Scientific Working Group on DNA Analysis Methods (SWGDAM). Guidelines for mitochondrial DNA (mtDNA) nucleotide sequence interpretation. *Forensic Sci. Commun.* [online] **5,** 2003.
88. Melton, T. and Nelson, K. Forensic mitochondrial DNA analysis: two years of commercial casework experience in the United States. *Croat. Med. J.* **42:**298–303, 2001.
89. Jobling, M. A. and Tyler-Smith, C. Fathers and sons: the Y chromosome and human evolution. *Trends Genet.* **11:**449–456, 1995.
90. Foster, E. A., Jobling, M. A., Taylor, P. G., et al. Jefferson fathered slave's last child. *Nature* **396:**27–28, 1998.
91. Ciminelli, B. M., Pompei, F., Malaspina, P., et al. Recurrent simple tandem repeat mutations during human Y-chromosome radiation in Caucasian subpopulations. *J. Mol. Evol.* **41:**966–973, 1995.
92. Kayser, M., Caglia, A., Corach, D., et al. Evaluation of Y-chromosomal STRs: a multicenter study. *Int. J. Legal Med.* **110:**125–133, 141–149, 1997.
93. Perez-Lezaun, A., Calafell, F., Seielstad, M., et al. Population genetics of Y-chromosome short tandem repeats in humans. *J. Mol. Evol.* **45:**265–270, 1997.
94. Prinz, M., Boll, K., Baum, H., and Shaler, B. Multiplexing of Y chromosome specific STRs and performance for mixed samples. *Forensic Sci. Int.* **85:**209–218, 1997.
95. Redd, A. J., Clifford, S. L., and Stoneking, M. Multiplex DNA typing of short-tandem repeat loci on the Y chromosome. *Biol. Chem.* **378:**923–927, 1997.
96. Prinz, M., Ishii, A., Coleman, A., Baum, H. J., and Shaler, R. C. Validation and casework application of a Y chromosome specific STR multiplex. *Forensic Sci. Int.* **120:**177–188, 2001.
97. Gonzalez-Neira A., Elmoznino, M., Lareu, M. V., et al. Sequence structure of 12 novel Y chromosome microsatellites and PCR amplification strategies. *Forensic Sci. Int.* **122:**19–26, 2001.
98. Bosch, E., Lee, A. C., Calafell, F., et al. High resolution Y chromosome typing: 19 STRs amplified in three multiplex reactions. *Forensic Sci. Int.* **125:**42–51, 2002.
99. Butler, J. M., Schoske, R., Vallone, P. M., Kline, M. C., Redd, A. J., and Hammer, M. F. A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. *Forensic Sci. Int.* **129:**10–24, 2002.
100. Sinha, S. K., Budowle, B., Arcot, S. S., et al. Development and validation of a multiplexed Y-chromosome STR genotyping system, Y-PLEX 6, for forensic casework. *J. Forensic Sci.* **48:**93–103, 2003.
101. Schoske, R., Vallone, P. M., Ruitberg, C. M., and Butler, J. M. Multiplex PCR design strategy used for the simultaneous amplification of 10 Y chromosome short tandem repeat (STR) loci. *Anal. Bioanal. Chem.* **375:**333–343, 2003.
102. Meng, C. Y., Both, K., and Burgoyne, L. A. Casework: A Y-STR triplex for use after autosomal multiplexes. *Forensic Sci. Commun.* [online] **5,** 2003.
103. Hall, A. and Ballantyne, J. The development of an 18-locus Y-STR system for forensic casework. *Anal. Bioanal. Chem.* [online], **376:** 1234–1246, 2003.
104. Sinha, S. K., Nasir, H., Gross, A. M., Budowle, B., and Shewale, J. G. Development and validation of the Y-PLEX 5, 1 Y-chromosome STR genotyping system, for forensic casework. *J. Forensic Sci.* **48:**985–1000, 2003.
105. Gusmao, L., Gonzalez-Neira, A., Pestoni, C., Brion, M., Lareu, M. V., and Carracedo, A. Robustness of the Y STRs DyS19, DYS389 I and II, DYS390 and DYS393: optimization of a PCR pentaplex. *Forensic Sci. Int.* **106:**163–172, 1999.
106. Carracedo, A., Beckmann, A., Bengs, et al. Results of a collaborative studey of the EDNAP group regarding the reproducibility and robustness of the Y-chromosome STRs DyS19, DYS389 I and II, DYS390, and DYS393 in a PCR pentaplex format. *Forensic Sci. Int.* **119:**28–41, 2001.
107. Jobling, M. A., Pandya A., and Tyler-Smith, C. The Y chromosome in forensic analysis and paternity testing. *Int. J. Legal Med.* **110:**118–124, 1997.
108. Prinz, M., Ishii, A., Coleman, A., Baum, H. J., and Shaler, R. C. Validation and casework application of a Y chromosome specific STR multiplex. *Forensic Sci. Int.* **120:**177–188, 2001.

109. Dekairelle, A. F. and Hoste, B. Application of a Y-STR-pentaplex PCR (DYS19, DYS389I and II, DYS390 and DYS393 to sexual assault cases. *Forensic Sci. Int.* **118:**122–125, 2001.
110. Gill, P., Brenner, C., Brinkmann, B., et al. DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs. *Forensic Sci. Int.* **124:**5–10, 2001.
111. National Institute of Justice. Solicitation for forensic DNA research and development for FY 2003. US Department of Justice, Office of Justice Programs. Washington, DC, 2003.
112. Lowe, A. L., Urquhart, A., Foreman, L. A., and Evett, I. W. Inferring ethnic origin by means of an STR profile. *Forensic Sci. Int.* **119:**17–22, 2001.
113. Grimes, E. A., Noake, P. J., Dixon, L., and Urquhart, A. Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype. *Forensic Sci. Int.* **122:**124–129, 2001.
114. Ayres, K. L., Chaseling, J., and Balding, D. J. Implications for DNA identification arising from an analysis of Australian forensic databases. *Forensic Sci. Int.* **129:**90–98, 2002.
115. Associated Press. DNA test showing ancestry may have helped La. Search. *USA Today*, June 5, 2003.
116. Schneider, P. M, Seo, Y., and Rittner, C. Forensic mtDNA hair analysis excludes a dog from having caused a traffic accident. *Int. J. Legal Med.* **112:**315–316, 1999.
117. Padar, Z., Angyal, M., Egyed, B., et al. Canine micosatellite polymorphisms as the resolution of an illegal animal death case in a Hungarian zoological gardens. *Int. J. Legal Med.* **115:**79–81, 2001.
118. Menotti-Raymond, M., David, V.A., Stephens, J. C., Lyons, L. A., and O'Brien, S. J. Genetic individualization of domestic cats using feline STR loci for forensic applications. *J. Forensic Sci.* **42:**1039–1051, 1997.
119. Padar, Z., Egyed, B., Kontadakis, K., et al. Canine STR analyses in forensic practice. Observation of a possible mutation in a dog hair. *Int. J. Legal Med.* **116:**286–288, 2002.
120. Butler, J. M., David, V., O'Brien, S., and Menotti-Raymond, M. The MeowPlex: a new DNA test using tetranucleotide STR markers for the domestic cat. *Profiles DNA* **5:**7–10, 2002.
121. Menotti-Raymond, M., David, V., Wachter, L., Yuhki, N., and O'Brien, S. J. Quantitative polymerase chain reaction-based assay for estimating DNA yield extracted from domestic cat specimens. *Croat. Med. J.* **44:**327–331, 2003.
122. Savolainen, P., Arvestad, L., and Lundeberg, J. A novel method for forensic DNA investigations: repeat-type sequence analysis of tandemly repeated mtDNA in domestic dog. *J. Forensic Sci.* **45:**990–999, 2000.
123. NDIS Statistics. Combined DNA Index System. June, 2003. Available at http://www.fbi.gov/hq/lab/codis/index1.htm.
124. Chanock, S. Candidate genes and single nucleotide polymorphisms (SNPs) in the study of human disease. *Dis. Markers* **17:**89–98, 2001.
125. Botstein, D. and Risch, N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat. Genet.* **33**(Suppl.)**:**228–237, 2003.
126. Jobling, M. A. Y-Chromosomal SNP haplotype diversity in forensic analysis. *Forensic Sci. Int.* **118:**158–162, 2001.
127. Wang, D. G., Fan, J. B., Siao, C. J., et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280:**1077–1082, 1998.
128. Heller, M. J. DNA microarray technology: devices. *Annu. Rev. Biomed. Eng.* **4:**129–153, 2002.
129. Linacre, A. and Graham, D. Role of molecular diagnostics in forensic science. *Expert Rev. Mol. Diagn.* **2:**346–353, 2002.
130. Salisbury, B. A., Pungliya, M., Choi, J. Y., Jiang, R., Sun, X. J., and Stephens, J. C. SNP and haplotype variation in the human genome. *Mutat. Res.* **526:**53–61, 2003.
131. Gill, P., Ivanov, P. L., Kimpton, C., et al. Identification of the remains of the Romanov family by DNA analysis. *Nat. Genet.* **6:**130–135, 1994.
132. Strom, C. M. and Rechitsky, S. Use of nested PCR to identify charred human remains and minute amounts of blood. *J. Forensic Sci.* **43:**696–700, 1998.
133. Schmerer, W.M., Hummel, S., and Herrmann, B. Optimized DNA extraction to improve reproducibility of short tandem repeat genotyping with highly degraded DNA as target. *Electrophoresis* **20:**1712–1716, 1999.
134. Gill, P., Whitaker, J., Flaxman, C., Brown, N., and Buckleton, J. An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. *Forensic Sci. Int.***112:**17–40, 2000.
135. Whitaker, J. P., Cotton, E. A., and Gill, P. A comparison of the characteristics of profiles produced with the AMPFlSTR SGM Plus multiplex system for both standard and low copy number (LCN) STR DNA analysis. *Forensic Sci. Int.* **123:**215–223, 2001.
136. Gill, P. Application of low copy number DNA profiling. *Croat. J. Med.* **42:**229–232, 2001.
137. Hawkins, T. L., Detter, J. C., and Richardson, P. M. Whole genome amplification—applications and advances. *Curr. Opin. Biotechnol.* **13:**65–67, 2002.
138. Telenius, H., Carter, N. P., Bebb, C. E., Nordenskjold, M., Ponder, B. A., and Tunnacliffe, A. Degenerate oligonucleotide-primed PCR: general amplification of target DAN by a single degenerate primer. *Genomics* **13:**718–725, 1992.
139. Barbaux, S., Poirier, O., and Cambien, F. Use of degenerate oligonucleotide primed PCR (DOP-PCR) for the genotyping of low-concentration DNA samples. *J. Mol. Med.* **79:**329–332, 2001.
140. Zhang, L., Cui X., Schmitt, K., Hubert, R., Navidi, W., and Arnheim, N. Whole genome amplification from a single cell: implications for genetic analysis. *Proc. Natl. Acad. Sci. USA* **89:**5847–5851, 1992.
141. Dean, F. B., Hosono, S., Fang, L., et al. Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. USA* **99:**5261–5266, 2002.
142. Cheung, V. G. and Nelson, S. F. Whole genome amplification using a degenerate oligonucleotide primer allows hundreds of genotypes to be performed on less than one nanogram of genomic DNA. *Proc. Natl. Acad. Sci. USA* **93:**14,676–14,679, 1996.

41 Molecular Assessment of Bone Marrow Transplant Engraftment

JESSICA K. BOOKER

1. INTRODUCTION

Allogeneic bone marrow transplantation has been used to treat a wide variety of inherited diseases, including hemoglobinopathies, immunodeficiencies, autoimmunities and lysosomal storage diseases, as well as acquired diseases such as hematologic and solid tumors *(1)*. Improvements in transplant protocols and the development of more effective immunosuppressive agents have both contributed to an increased use of allogeneic bone marrow transplantation. Advances in the methodology used to monitor bone marrow transplant engraftment status have contributed to the clinical utility of the information provided by molecular assessment. Turnaround time for molecular engraftment studies has decreased from weeks to days, providing clinicians with the opportunity to include the results into their decisions regarding treatment. In addition to rapid availability of results, quantitative measurements can be used to evaluate graft status at any given time and to monitor changes over time. Increased sensitivity of molecular engraftment studies can provide the earliest indication of graft rejection or disease recurrence, allowing for faster clinical intervention using means such as donor lymphocyte infusion, immunosuppressive agent modification, additional drug or radiation therapy, and consideration of a second transplant *(1)*.

There are many types of allogeneic bone marrow transplant that can be performed. Variables include the cells used (bone marrow, cord blood, peripheral blood), treatment protocols on the cells prior to transplant (T-cell depletion), the relationship of the donor to the recipient (matched sibling, matched unrelated, partially matched related, cord blood), and the conditioning regimen the recipient undergoes prior to receiving the transplant (myeloablative, nonmyeloablative) *(2)*. Regardless of the combination of variables employed, the molecular methods used for monitoring posttransplant samples are the same, although the clinical interpretation of the results might differ. For example, engraftment kinetics is expected to be different in a patient who received a myeloablative pretransplant regimen compared to one that underwent a reduced intensity conditioning regimen. In addition, engraftment kinetics might differ between cells of myeloid and lymphoid lineage. Molecular assessment of the engraftment status of specific lineages is available clinically and is used as an early indication of graft-vs-host disease.

2. MONITORING BONE MARROW TRANSPLANT ENGRAFTMENT

Once a bone marrow transplant has been performed, the recipient is monitored to evaluate the status of the graft. Clinical decisions regarding immunosuppressive therapy, chemotherapy, radiation therapy, and donor lymphocyte infusions are determined by a combination of indicators, including the results of engraftment studies.

2.1. UTILITY OF MONITORING ENGRAFTMENT

Molecular assessment of bone marrow transplant engraftment is used for the monitoring of graft rejection, minimal residual disease, recurrence, and graft-vs-host disease. Monitoring engraftment status over time provides information regarding current graft status, trends, and response to therapies. Monitoring can be done on posttransplant samples from bone marrow or peripheral blood, with comparable results obtained under most conditions. Analysis of bone marrow might be preferable when the purpose is to detect minimal residual disease *(1)*.

2.2. METHODS Methods used for identity testing rely on DNA polymorphisms that occur at a high enough frequency to be useful in distinguishing one individual from another. Monitoring bone marrow transplant engraftment utilizes the same methods as those used in forensics and paternity testing. Methodology has generally shifted from restriction fragment length polymorphism (RFLP) *(3)* to variable number tandem repeat (VNTR) *(4)* to short tandem repeats (STRs) *(4)*, and from detection by Southern blotting to traditional gel electrophoresis to capillary gel electrophoresis. The most widely used method at this time is the analysis of multiple STR loci by polymerase chain reaction (PCR) with detection by capillary gel electrophoresis. STRs, also known as microsatellites, are spread throughout the genome and consist of repeats 2–7

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

nucleotides long, repeated multiple times in tandem array. The range of repeats at any given STR loci is low enough that there is little risk of a bias for amplifying a smaller-sized allele over a larger one. The application of capillary gel electrophoresis for the detection of microsatellite markers has dramatically impacted the turnaround time. Capillary gel electrophoresis allows for quantification of the contributions from multiple genotypes in a chimeric sample. Kits are available for identity testing with fluorescently labeled primer pairs that are either sold individually or multiplexed for the concurrent amplification of up to 16 microsatellite markers in a single tube *(5)*. Primers are designed to flank the repeat region so that the length of the PCR product is determined by the number of repeats *(6)*. An additional advantage of the more recent methods such as STR analysis using capillary electrophoresis or single-nucleotide polymorphism (SNP) analysis is the minimal amount of DNA required, particularly when compared to Southern blot analysis. Generally, 1–10 ng of DNA yields optimal results *(7)*.

The future direction of identity testing might be the use of quantitative SNP analysis by real-time PCR, which would eliminate any issues involved in the amplification of nucleotide repeat sequences *(8)*.

2.3. INFORMATIVE MARKERS Under optimal conditions, donor and pretransplant recipient samples are available and they are used to identify informative markers for optimal distinction between donor and recipient cells in the bone marrow and peripheral blood of subsequent posttransplant samples. The objective is not only to be able to find markers that distinguish donor from recipient but also those that allow for the detection of low levels of recipient cells in a population of cells that is predominantly of donor origin.

A complicating artifact resulting from amplification of STRs is known as "stutter" *(9)*. When the polymerase is amplifying through extended regions of repeats, there is a slight tendency to slip, the end result being an amplicon one repeat shorter than the full length. Given a great enough propensity for stutter and enough amplification cycles, a single true amplicon can give rise to a stutter amplicon one repeat shorter than the true allele, and this stutter amplicon can, in turn, give rise to its own stutter amplicon. Stutter amplicons resulting from amplification of tetranucleotide repeats usually represent about 5% of the normal amplicon *(10)*. The interference resulting from stutter can be minimized by the choice of STRs in that the smaller the number of nucleotides in a repeat, the greater the likelihood of polymerase slippage, which is why the microsatellite markers used in identity testing are generally tetranucleotide repeats rather than dinucleotide or trinucleotide repeats. When selecting informative markers, the potential for stutter and the impact on interpretation must be taken into consideration. Take for example an STR marker for which the donor has 1 allele with 12 repeats and the other with 10 repeats and the recipient has 1 allele with 12 repeats and the other with 9 repeats. Donor and recipient both have an allele with 12 repeats, so this is not informative, but they differ at their other allele, which makes this STR marker informative in that it can be used to distinguish between these two individuals. If, however, this marker is used to monitor the patient posttransplant for the presence of a minority of cells of recipient origin and the 9-repeat allele is detected, it will be impossible to determine whether this represents cells of recipient origin or stutter from the 10-repeat allele of the donor. In this example, even though the donor and recipient differ at one allele of this marker, it would not be informative for identifying small numbers of cells of recipient origin and is therefore not a good choice.

The most efficient way to identify informative markers, particularly for related donor/recipient pairs, is to use a multiplex kit. In this way, each individual can be genotyped for up to 16 microsatellite markers at one time, with a high likelihood of identifying multiple informative markers. Once informative markers have been identified from the pretransplant and recipient samples, posttransplant samples can be tested either with the same multiplex kit or with the individual markers that were found to be informative.

2.3.1. Number of Many Informative Markers to Use Determination of the number of markers to use partially depends on the number of informative markers that are available. A bone marrow transplant from an unrelated donor generally results in multiple informative markers that can be used to differentiate between cells of donor and recipient origin in posttransplant studies. A significant proportion of bone marrow transplants are between close relatives, which can make the identification of informative markers a challenge. When multiple markers are available and posttransplant samples are analyzed by a multiplex kit, all available markers can be used and quantitative results can be averaged. When multiple markers are available and posttransplant samples are going to be analyzed by monoplex markers, two to three markers are adequate, particularly if each is located on a different chromosome. If only a single informative marker is available, it can be used, but the interpretation must be made cautiously. With a single informative marker, there is a possibility that a chromosomal abnormality in the recipient marrow can interfere with accurate interpretation of quantitative results. Considering that bone marrow transplantation is used to treat hematologic malignancies, and malignant cells frequently exhibit genetic instability, it is possible that a recipient population (tumor) could recur and that it could have a chromosomal deletion or duplication. If a single marker is used for engraftment studies and the marker is on the same chromosome that has been deleted or duplicated, quantitative results will be inaccurate, at best, and entirely misleading, at worst *(11)*.

2.3.2. Less Than Ideal Circumstances Under ideal circumstances, the laboratory conducting posttransplant bone marrow engraftment studies has access to donor and pretransplant recipient samples for identification of informative markers and measurements of sensitivity when posttransplant samples are tested. It is not uncommon for one or both of these samples (donor or pretransplant recipient) to be unavailable to the laboratory performing the posttransplant engraftment analysis. When a pretransplant recipient sample is unavailable, a buccal swab can be obtained from the recipient, anytime after the transplant. In theory, cells from a buccal swab are epithelial and should be of recipient origin *(12)*. In practice, perhaps because of the immunosuppressive agents that bone marrow

transplant recipients receive, there is often a significant contribution of donor cells in the buccal swab. When mouthwash is used to collect the sample, the cells collected can be entirely of donor origin *(13)*. The presumed pretransplant recipient genotype can be identified by comparing the genotype of the cells in the buccal swab (which will either be pure recipient, or a chimera of recipient and donor) with the donor sample. If a donor sample is unavailable, the donor genotype can be presumed from a comparison of the pretransplant recipient with a posttransplant sample (if the recipient is doing well, the posttransplant sample will either be pure donor or a chimera of recipient and donor). Unavailability of both donor and pretransplant recipients really puts the interpreter of fingerprinting results at a disadvantage. Even under these adverse conditions, it is still potentially possible to presume donor and pretransplant recipient genotypes from a comparison of posttransplant marrow or peripheral blood and buccal swab samples from the recipient. Presumption of donor and/or recipient genotypes must be interpreted with care and the potential limitations should be indicated in the clinical report. When either a donor or pretransplant recipient sample is unavailable, a sensitivity control for that particular pair cannot be used.

2.4. QUANTIFICATION Quantitative fingerprinting results provide valuable information to the clinician and are particularly informative when engraftment status is monitored quantitatively over time. The ability to detect graft rejection, graft-vs-host disease and recurrence can potentially be detected earlier and the efficacy of therapy given to support engraftment can be monitored.

2.4.1. Calculating Donor and Recipient Contributions in Chimeric Populations Capillary gel electrophoresis of amplified STR loci is the method most commonly used for quantitative engraftment studies. Once the donor and recipient pretransplant genotypes have been determined for the chosen loci, posttransplant samples can be analyzed to identify the donor and recipient contributions if the posttransplant sample is chimeric. Figure 1 shows examples of the calculations used for quantitative results. When more than one informative marker is used, the appropriate calculation is done for each individual marker and the results can be averaged.

2.5. SENSITIVITY Different laboratories are able to achieve varying degrees of sensitivity depending on the method used. The ability to detect a contribution of 0.1–5% of recipient cells in a background of 95–99.9% of donor cells is typical, and generally meets the needs of the clinicians *(1)*. The sensitivity that can be reproducibly attained must be determined for the system employed and should be demonstrated every time the assay is performed. Approaches to controlling for reported sensitivity is addressed in Section 2.6. When very low levels of cells of recipient origin are present, they might be below the level of detection. Sensitivity under these circumstances can be increased by using lineage-specific analysis in which cells of myeloid and lymphoid lineage are separated prior to analysis *(1)*.

2.6. CONTROLS Controls must be included every time a sample is run to ensure that the assay is performing as expected. For every PCR reaction, a setup is done with all the reagents used for amplification, with water used in place of DNA to demonstrate that none of the reagents are contaminated. In addition, it is necessary to control for the reported level of sensitivity. The sensitivity must be determined for the system employed and then this sensitivity must be demonstrated with every run. One way of demonstrating the reported sensitivity with each assay is by using a standard control that is made by mixing DNA from two unrelated individuals at the appropriate concentration that corresponds to the reported level of sensitivity (i.e., 1%, 5%, or 10%). When the sensitivity control is included with each run, detection of the minor population indicates that the assay is sensitive to the reported degree.

There is variability in the behavior between different microsatellite markers and there is variability in the behavior of different combinations of alleles within each microsatellite marker. For these reasons, the most conservative approach to controlling for sensitivity is to run a control for each donor and recipient pair, for each informative marker used, with each run. When cells of recipient origin are not detected in a posttransplant sample but are detected in the control (say a 5% mix of recipient cells in a 95% background of donor cells), then the results can be reported as negative for the presence of recipient cells to a sensitivity of 5%. If cells of recipient origin are undetected in both the posttransplant and sensitivity control samples, the assay must be repeated, as a level of sensitivity cannot be reported. If cells of recipient origin are undetected in the sensitivity control but are detected in the posttransplant sample, it is legitimate to report those results, as the purpose of the sensitivity control is to be able to report the level of sensitivity when negative results are obtained. It is important that levels of amplification be comparable between the posttransplant and control samples, whether detection is by Southern blotting, acrylamide gel, or capillary gel electrophoresis. If amplification of the predominant allele in a posttransplant sample is weak, relative to the predominant allele in the control, and cells of recipient origin are undetected, it is not accurate to report the level of sensitivity attained in the control. In circumstances such as this, an attempt should be made to improve amplification from the posttransplant sample.

2.7. REPORTING Reporting of engraftment studies should include the method, the markers used, and the sensitivity of the assay. The average quantitative result should be reported along with the range, to give an indication of the variability. Any limitations of the assay, including absence of donor or pretransplant recipient samples resulting in the necessity to presume those genotypes, should be mentioned. In the case of a single informative marker used for reporting, the limitations should be addressed.

2.8. CONSIDERATION OF BONE MARROW TRANSPLANTS IN OTHER MOLECULAR TESTING With the increasing use and success of bone marrow transplantation, there are an increasing number of chimeric individuals in the population. When other kinds of genetic tests are performed on blood samples from these patients, it is important for the origin of the sample to be considered in the interpretation of the results.

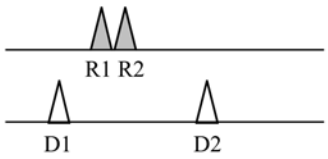
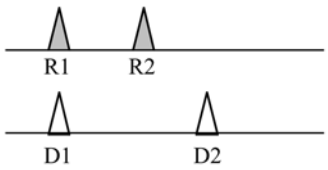
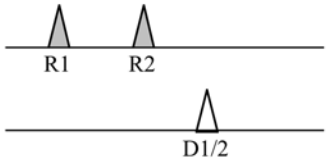
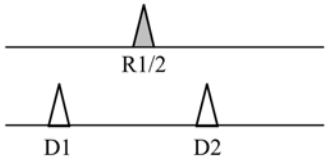
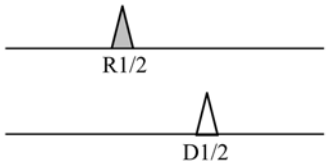
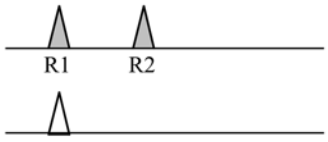
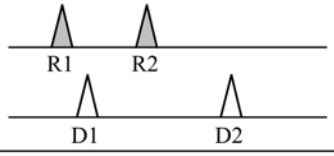
Donor and Recipient Alleles	Calculation of % Recipient	Comment
	$\frac{R1+R2}{R1+R2+D1+D2}$	2 informative alleles
	$\frac{R2}{R2+D2}$	1 informative allele
	$\frac{R1+R2}{R1+R2+D1/2}$	2 informative alleles
	$\frac{R1/2}{R1/2+D1+D2}$	2 informative alleles
	$\frac{R1/2}{R1/2+D1/2}$	2 informative alleles
	$\frac{2(R2)}{R2+D1/2}$	1 informative allele
	$\frac{R2}{R2+D2}$	1 informative allele 1 potential stutter

Fig. 1. Calculations for informative STR alleles. R1 and R2 are recipient alleles and R1/2 is a homozygous recipient allele. D1 and D2 are donor alleles, and D1/2 is a homozygous donor allele. For STR analysis by capillary gel electrophoresis, calculations are done using the peak area. (From ref. *7*.)

REFERENCES

1. Antin, J. H., Childs, R., Filipovich, A. H., et al. Establishment of complete and mixed donor chimerism after allogeneic lymphohematopoietic transplantation: recommendations from a workshop at the 2001 Tandem Meetings. *Biol. Blood Marrow Transplant.* **7:**473–485, 2001.
2. Applebaum, F. R. The current status of hematopoietic cell transplantation. *Ann. Rev. Med.* **54:**491–512, 2003.
3. Ginsburg, D., Antin, J. H., Smith, B. R., Orkin, S. H., and Rappeport, J. M. Origin of cell populations after bone marrow transplantation. Analysis using DNA sequence polymorphisms. *J. Clin. Invest.* **75:**596–603, 1985.
4. Koreth, J, O'Leary, J. H., and McGee, J. Microsatellites and PCR genomic analysis. *J. Pathol.* **178:**239–248, 1996.
5. Nollet, F., Billiet, J., Selleslag, D., and Criel, A. Standardisation of multiplex fluorescent short tandem repeat analysis for chimerism testing. *Bone Marrow Transplant.* **28:**511–518, 2001.
6. Jeffreys, A., Wilson, V., Neumann, R., and Keyte, J. Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells. *Nucleic Acids Res.* **16:**10953–10971, 1988.

7. Van Deerlin, V. M. and Leonard, D. G. Bone marrow engraftment analysis after allogeneic bone marrow transplantation. *Clin. Lab. Med.* **20:**197–225, 2000.
8. Oliver, D. H., Thompson, R. E., Griffin, C. A., and Eshleman, J. R. Use of single nucleotide polymorphisms (SNP) and real-time polymerase chain reaction for bone marrow engraftment analysis. *J. Mol. Diagn.* **2:**202–208, 2000.
9. Walsh, P. S., Fildes, N. J., and Reynolds, R. Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. *Nucleic Acids Res.* **24:**2807–2812, 1996.
10. Schichman, S. A., Suess, P., Vertino, A. M., and Gray, P. S. Comparison of short tandem repeat and variable number tandem repeat genetic markers for quantitative determination of allogeneic bone marrow transplant engraftment. *Bone Marrow Transplant.* **29:**243–248, 2002.
11. Schichman, S. A., Lin, P., Gilbrech, L. J., Gray, P. S., Wilson, C. S., and Sawyer, J. R. Bone marrow transplant engraftment analysis with loss of an informative allele. *J. Mol. Diagn.* **4:**230–232, 2002.
12. Thiede, C., Prange, Krex, G., Freiberg-Richter, J., Bornhauser, M., and Ehninger, G. Buccal swabs but not mouthwash samples can be used to obtain pretransplant DNA fingerprints from recipients of allogeneic bone marrow transplantation. *Bone Marrow Transplant.* **25:**575–577, 2000.
13. Endler, G., Greinix, H., Winkler, K., Mitterbauer, G., and Mannhalter, C. Genetic fingerprinting in mouthwashes of patients after allogeneic bone marrow transplantation. *Bone Marrow Transplant.* **24:**95–98, 1999.

42 Use of DNA-Based Identity Testing for Specimen Identification

ANDREW RICCI JR. AND GREGORY J. TSONGALIS

1. INTRODUCTION

Both state and federal regulations mandate the administration of a Quality Assurance program for all clinical laboratories. As one of the preanalytical testing variables, all clinical specimens must be followed by documentation that includes patient demographic data, a unique identifier such as an accession number, and date/time of collection. Despite various quality assurance efforts such as these, specimen mislabeling or mix-ups continue to be potential problems for the clinical laboratory. Serious clinical consequences could result from the inability to accurately resolve such issues. Nucleic acid analysis for the diagnosis of disease has become a routine means of testing in the clinical laboratory. It is now also widely accepted that DNA typing can either associate or exclude an individual as the source of a given biological sample.

Identity testing in criminal cases began with the introduction of traditional fingerprinting first described in the 1800s. ABO blood group typing was introduced in 1900 and human leukocyte antigen (HLA) tissue typing followed in the 1960s *(1–3)*. More revealing techniques have now evolved from our understanding of the biochemical properties of deoxyribonucleic acid (DNA), which have had a major impact on the forensic sciences for the purpose of establishing identity. In 1980, Wyman and White were the first to describe a hypervariable region of DNA consisting of variable numbers of tandem repeats (VNTRs) *(4)*. These interindividual genetic differences result from inconsequential changes or polymorphisms in untranscribed DNA sequences that are not usually associated with disease. Our ability to detect these variations is the foundation for DNA-based identity testing. Several years after the elucidation of VNTRs, the first DNA "fingerprint" was described and entered into court *(5–8)*. Most recently, our knowledge of the human genome has led to an explosion of new DNA-based methods for identity testing requiring minimal amounts of tissue or body fluid for analysis *(9–26)*. It is evident that DNA analysis for identity purposes can play an important role in specimen identification in the clinical laboratory. This chapter will present an overview of this unique application and discuss several clinical cases in which it was successfully employed.

2. DNA STRUCTURE

In 1953, Watson and Crick wrote, "This [DNA] structure has novel features which are of considerable biological interest" *(27)*. In that same year, they expanded on the chemical composition of deoxyribonucleic acid (DNA) and elucidated its role as the genetic material that is now recognized as the "blueprint" of all living things *(28)*. DNA consists of a double helix bound by complementary base-pairing of repetitive coding and noncoding nucleotides. It is estimated that the human genome is comprised of between 30,000 and 60,000 genes compactly packaged into 46 chromosomes consisting of 22 pairs of autosomes and 1 pair of sex chromosomes (X or Y). A haploid genome (23 chromosomes) containing only half of the total genetic material, as in sperm and egg cells (gametes), contains approx 3×10^9 basepairs (bp). The majority of human DNA exists within the nucleus of the cell, but human DNA also resides outside of the nucleus within the cell's mitochondria. Of the 3×10^9 bp of genomic DNA, only 3–5% account for coding sequences (exons) that result in translated protein. The remaining 95–97% consists of intervening sequences (introns). Seventy-five percent of introns consist of unique base-pair sequences, whereas 25% are repetitive *(29)*.

An interesting aspect of the noncoding repetitive sequences, fundamental to DNA-based identity testing, is the amount of interindividual variation that these sequences exhibit. In some instances, these variants are the result of mutation, the permanent alteration of DNA sequences. However, when two or more variants are present in a given population with a frequency of greater than 1%, they are said to be "polymorphic." At the molecular level, DNA polymorphism refers to differences in nucleotide sequences (alleles) between two chromosomes at the same locus. Genetic polymorphism most often is a normal occurrence without negative consequences and actually results in those traits that establish our individuality. The first types of polymorphism to be described were those that resulted in fragment length variations that are detectable when DNA is

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

digested with restriction endonucleases *(30,31)*. These enzymes cleave DNA at specific recognition sites so that smaller DNA fragments are produced. Restriction fragment length polymorphisms (RFLP) are apparent when a restriction endonuclease recognition sequence varies between alleles at the same locus. RFLPs resulting from such allelic variation can be detected within the same individual or between individuals. These are commonly referred to as sequence polymorphisms. A second type of polymorphism that is seen is based on sequence length or size. Size polymorphisms consist of hypervariable regions of DNA and are characterized by repetitive units of the same DNA sequence. Minisatellite sequences were the first of these types of polymorphism to be identified. These are the so-called variable numbers of tandem repeats (VNTRs), which can be from 10 to 100 bp in length *(5)*. Microsatellites or small tandem repeats (STRs) are also seen, and they consist of dinucleotide, trinucleotide, and tetranucleotide repetitive sequences that are distributed throughout the genome with a frequency of approx 1 locus per 10 kb of sequence *(32)*.

3. TECHNOLOGY

3.1. SPECIMENS DNA can be isolated from most clinical specimens, including very small paraffin-embedded tissue biopsies and even histologic tissue sections from glass slides. Depending on the specimen type and size, various extraction protocols can be employed. Nonorganic extraction methods are preferred and can be performed manually or on automated platforms. Most body fluids can be performed on automated platforms. Tissue specimens, however, require more manipulation and, thus, extractions are best performed manually. This can be aided by current magnetic particle technologies that can, in addition to extraction, also be used to simultaneously semiquantitate DNA, as only a finite amount of nucleic acid will bind to the coated particles.

DNA extraction of appropriate tissues and/or cell types is critical to the identification process. For mislabeled specimens, it is important to have documentation concerning the exact scenario that took place in the handling of the specimen (i.e., why was specimen identification even a question?). With respect to histological material, analysis should begin with the review of materials by an Anatomic Pathologist, who can (1) verify the nature of the submitted tissue and (2) demarcate relevant areas of tissue for DNA extraction. Simple procedures are currently available to select and demarcate tissues in paraffin-embedded blocks or from stained or unstained tissue sections on glass slides. "Floaters" (tissue derived from extraneous sources) and/or other relevant tissues on a glass slide can be marked with a diamond etching pen. Similarly, faced paraffin blocks can be highlighted with a permanent marking pen. The most suitable tissues for microdissection and DNA extraction should be cellular, nucleated, and non-necrotic. In order to identify a particular individual as the most likely source of a tissue sample, a comparative analysis on DNA extracted from a submitted whole-blood specimen or other verifiable tissue sample is required.

Diamond etching pens and permanent ink markers are relatively blunt instruments, but in many cases of identity testing, they are all that is needed for highlighting relevant tissues to be assayed. When greater precision is required, more refined and specific microdissection techniques that are capable of dissecting one cell from a complex tissue section can be employed. From crude dissecting microscope procedures, to use of chemical extractions and laser capture microdissection (LCM), the goal has always been the same: to obtain the correct specimen for analysis *(33, 34)*. All of the advanced technologies we could apply would come to nothing or even produce misleading results if the correct tissues were not being analyzed.

The PinPoint Slide DNA Isolation Sytem™ (Zymo Research, Orange, CA) utilizes a proprietary solution that can be applied to paraffin-embedded tissue sections on a glass slide *(33)*. The solution can be applied macroscopically or microscopically. Once dried, a scalpel blade is used to remove the isolation solution and adhered target tissue for DNA extraction.

Laser capture microdissection is a refined tool that permits reliable procurement of pure cell populations from tissue sections *(34)*. Briefly, a histologic section is placed on the stage of a specialized microscope and an area of interest is brought into view. A cap coated with a thermoplastic film is then placed over the cells to be dissected and a low-power infrared laser beam is activated. The laser melts the thermoplastic film that then binds to the cells beneath it. In this way, LCM can remove large clusters of cells or single cells as a pure population, free from supporting stromal or other contaminating elements.

3.2. SOUTHERN BLOT A significant advance in nucleic acid biology came in 1975 when Edward M. Southern described what has come to be known as the Southern blot method for analyzing DNA *(35)*. In this method, DNA is extracted from the specimen and digested with specific restriction endonucleases. Once the isolated DNA has been cleaved by the endonucleases, the DNA fragments are separated by gel electrophoresis and transferred to a solid support (nitrocellulose or nylon membrane). Subsequent hybridization with sequence-specific probes allows for the visualization of banding patterns with either radioactive, colorimetric, or chemiluminescent detection systems. The Southern blot was originally used to detect RFLPs. Using this technique, different banding patterns resulted reflecting which restriction endonuclease recognition sequences were present or absent. The discovery of hypervariable sequences subsequently allowed for Southern blot analysis of multiple fragments based on size variation using probes for specific repetitive sequences *(4,5)*.

3.3. POLYMERASE CHAIN REACTION While Southern blot transfer technologies became recognized by the scientific community and accepted by the courts as a valuable tool for further study of the genome and for identity testing, a second technological advancement was made. This entailed the capability to synthesize complementary DNA sequences in vitro with concomitant amplification of target DNA sequences *(36–38)*. This methodology, known as polymerase chain reaction (PCR), requires the use of short oligonucleotides (primers) that hybridize to their complementary sequences and thus identify the target sequence for amplification. These sites are then the start sites for synthesis of new DNA strands by thermostable polymerases. DNA is first denatured at high temperatures and primer sequences are allowed to anneal to their

complementary sequences at lower temperatures. Subsequently, a polymerase will begin synthesis of new DNA strands. These three steps of denaturation, annealing, and synthesis or elongation constitute one PCR cycle *(38,39)*. Because this is a logarithmic amplification process, one million copies of a single target sequence are produced after only 20 cycles.

Polymerase chain reaction, the capability to synthesize complementary DNA sequences in vitro with concomitant amplification of target DNA sequences, has allowed us to interrogate nucleic acids extracted from very small amounts of fluid or tissue. PCR has proven quite successful in its ability to distinguish between variously sized polymorphisms simply by amplifying smaller or larger fragments depending on the number of repeats present *(13–26)*. The major advantages of PCR over other traditional molecular techniques such as the Southern blot are time/labor, assay sensitivity, ability to interrogate increasingly smaller specimens, and automation.

There are currently several automated DNA sequencing systems commercially available and each is suitable for fragment size analysis of PCR-amplified microsatellite sequences (STRs) *(40–43)*. These systems, such as the SGM Plus (Perkin Elmer, Foster City, CA) and the PowerPlex 16 (Promega, Madison, WI), allow for the coamplification of up to 16 loci using a 3-color fluorescent detection system. A once commonly used procedure for typing multiple loci simultaneously was the reverse dot blot procedure that uses an array of immobilized probes to detect sequence polymorphisms. The AmpliType PM + DQA1 amplification and typing kit (Roche Molecular Systems, Indianapolis, IN) was used routinely by many forensic laboratories *(13,20)*. This procedure had been applied to numerous cases of clinical specimen identification and has been able to produce genotype results from many different specimens *(33,44–46)*. Extremely reproducible, this kit interrogated six polymorphic loci in a single multiplex PCR assay.

4. CLINICAL UTILITY

DNA analysis has gained tremendous support with respect to the clinical laboratory through a multitude of diagnostic applications. Bone marrow engraftment analysis is an ideal example of applying identity testing in a clinical setting. Using technologies and polymorphic markers similar to forensic identity and parentage testing, these analyses can accurately assess the presence or absence of recipient vs donor hematopoietic cells. The ability to distinguish between two individuals using polymorphic DNA sequences is the essence of identity testing and can also be applied to specimen identification issues in the clinical laboratory *(33,44,47–49)*. The accurate identification of clinical specimens is essential to proper management of a patient and misidentification could lead to catastrophic overtreatment or withholding of necessary therapies.

Given the advances in molecular technologies with respect to DNA extraction procedures, in vitro amplification, and detection systems, it should come as no surprise that the challenge of positive specimen identification can be met. In most instances, we can accurately determine the identity or origin of a clinical specimen. Mislabeled or unlabeled clinical specimens are the most common identification issues for the clinical laboratory. A typical scenario might be a busy emergency room where blood samples are being drawn from several individuals at the same time. As a matter of protocol, a clinical laboratory will not assay improperly labeled or unlabeled blood specimens unless a firm chain of possession can be documented by the clinician. Ordinarily, blood samples would be redrawn. In the anatomic pathology laboratory, the issue of proper specimen identification is much the same; however, specimen recollection via rebiopsy might be difficult or impossible. Similar to blood specimens, tissue biopsies can be mislabeled at the time of collection (i.e., multiple biopsies from the same or different patients) or processing (tissue holders or glass slides in the laboratory). In addition to the potential mislabeling of specimens, physicians, might on occasion, be challenged by a patient who is in denial of their condition and does not believe that the tissue diagnosis could pertain to their medical condition. DNA identity testing can then be called upon to certify the origin of biologic specimens in all of these scenarios.

An uncommon but critical challenge for the anatomic pathologist is the interpretation of fragments of tissue known as "floaters" that histologically do not appear to be associated with the case at hand. This is not a labeling error or specimen mix-up but rather results from small tissue fragments being "carried over" from one case to the next. This represents a form of specimen contamination. These "floaters" can be carried over during gross dissection or later during histologic processing of tissue biopsies. This form of contamination can occur despite rigorous quality control procedures. In the majority of these cases, the anatomic pathologist is able to dismiss the questionable fragment of tissue as not germane to the case during routine histologic examination. This is easily accomplished when the "floater" does not resemble the histology of the remaining tissue, is positively recognized as coming from another case, or presents incongruous histology. However, when the "floater" is the same type of tissue as the case at hand, it becomes more difficult to simply discard it as irrelevant. In addition, "floaters" can represent just a few cells or clusters of cells that are difficult to classify and evaluate definitively. One of the more common scenarios we encounter in consultation is a case with a malignant "potential floater" that is only focally represented in the tissue section. Such tissue cannot easily be ignored, but neither can it be positively interpreted and reported. In the past, when diagnostic certainty could not be applied because of potential "floater" contamination, clinical recommendations were reduced to ". . . advise close follow-up and short term re-biopsy if possible. . . "

5. INTERPRETATION

Microsatellite instability (MSI) is a feature of some common tumor systems. For example, approx 20% of sporadic colon cancers are characterized as having a mutator phenotype that results in MSI. Because identity testing relies on polymorphic STR sequences, MSI could present a confounding influence in these assays. This is not the case with sequence polymorphisms as in the PM + DQA1 kit. It is important for an identity testing laboratory to document the genotypic conservation of the STRs used in their assays by testing of normal and

malignant tissues. The use of similar markers for LOH or MSI studies are not recommended in the analysis of malignant vs normal tissues, "floaters" or otherwise.

6. CASE REPORTS

The following case reports exemplify some of the issues associated with specimen identification testing. As with all tissue cases sent to us for identity testing, histopathologic review is performed before any testing is begun. This is done to verify the nature of the submitted material and to select appropriate tissues for comparative analysis.

6.1. CASE 1 This case represents an example of questioned but correct identity of two routine histologic hematoxylin and eosin (H&E) stained slides. In this case, we received two H&E-stained slides provisionally identified to us as tissues that originated from patient A. DNA was extracted from two H&E-stained slides prepared from paraffin-embedded tissue blocks. After histopathologic review, relevant tissues were microdissected and assayed. In addition, DNA was extracted from a whole-blood specimen identified to us as originating from patient A.

The histopathologic review indicated that the two slides (1 and 2) provisionally identified as patient A showed segments of fallopian tubes (two) showing normal histology, including complete cross sections. We also received an EDTA anticoagulated whole-blood specimen from the patient.

Marker	*Slide 1*	*Slide 2*	*Whole blood*
D3S1358	16,18	16,18	16,18
VWA	16,17	16,17	16,17
D16S539	11,12	11,12	11,12
D2S1338	19,25	19,25	19,25
D8S1179	13	13	13
D21S11	29,32.2	29,32.2	29,32.2
D18S51	13	13	13
D19S433	15,16.2	15,16.2	15,16.2
THO1	8	8	8
FGA	20,22	20,22	20,22

DNA typing was performed using a PCR-mediated STR assay (SGM Plus; Perkin-Elmer, Foster City, CA) analyzed by capillary electrophoresis. The results of this assay indicate that all three of these specimens most likely originated from the same source. The clinical issue might be obvious to the reader. The patient became pregnant after bilateral tubal ligation and wanted to ascertain that she received the correct surgical procedure.

6.2. CASE 2 This case represents an example of a questionable "floater" carried over from another case. We received two H&E-stained slides. DNA was extracted from three tissue fragments that were microdissected from these two H&E-stained slides.

The histopathologic review identified slide 1, representing the first patient, as a prostate core biopsy consisting of benign prostate tissue with a focus of transitional cell carcinoma (a suspected contaminating "floater" from another tissue source). In this case, the questionable "floater" (labeled A) and a fragment of normal prostate (labeled B) were microdissected from the same slide. Slide 2, representing a second patient, was shown to consist of a papillary transitional cell carcinoma and microdissected (labeled C).

System	*A (?floater)*	*B*	*C*
D3S1358	17,18	16	17,18
THO1	6,9.3	7,10	6,9.3
D21S11	31.2	29	31.2
D18S51	12,15	13,14	12,15
PENTA E	10,12	17,20	10,12
D5S818	11,12	11,12	11,12
D13S317	11	8,10	11
D7S820	9,11	12	9,11
D16S539	11	11,14	11
CSF1	12,13	10,12	12,13
PENTA D	9	12,15	9
VWA	15,18	17,19	15,18
D8S1179	10	12,14	10
TPOX	8,9	8,11	8,9
FGA	21,23	18,24	21,23
Amelogenin	X	XY	X

Using the PowerPlex 16 System (Promega Corp., Madison, WI), the results of this assay indicated that the questionable fragment of tissue (A) originated from the same source as the tissue on slide 2 (C). In comparison, the genotype from prostate tissue from slide 1 (B) did not match that from the other tissues. When combined with the likelihood of contamination based on histologic examination, the exclusion of identity (A vs B) is a virtual certainty. Further, in this case, the questionable fragment of tissue actually originated from a female (see amelogenin marker).

6.3. CASE 3 This case represents a potential mislabeling of surgical specimens. We received four H&E-stained slides for DNA extraction and typing analysis. After histopathologic review, tissues from these four H&E-stained slides were microdissected.

Histopathologic review showed slide 1 to be a right upper lung core biopsy consisting of a moderately differentiated partially necrotic clear cell adenocarcinoma and provisionally identified to us as patient A. Slide 2 was a right upper lobectomy consisting of a necrotizing granulomatous inflammation without the presence of carcinoma and also identified to us as patient A. Slide 3 was a left upper quadrant core biopsy of the abdomen showing chronic inflammation and granulation tissue without the presence of neoplasia and provisionally identified as patient B. Slide 4 was an endometrial biopsy showing a high-grade endometrioid carcinoma.

Marker	*1*	*2*	*3*	*4*
D3S1358	16	16,17	16,17	16
VWA	16,17	15,16	15,16	16,17
D8S1179	12,14	11,13	11,13	12,14
D21S11	28	27,30	27,30	28
D19S433	11,14	13,14	13,14	11,14
THO1	8	6,7	6,7	8
FGA	22	25	25	22
Amelogenin	X	XY	XY	X

The results of this analysis indicate that the tissue on slide 1 originated from the same individual as the tissue on slide 4. In addition, the tissue on slide 2 originated from the same individual as the tissue on slide 3. Also, note that tissues on slides 2 and 3 originated from a male source. The most likely explanation for these data is that the specimen designations for slides 1 and 3 had been reversed within the laboratory, as both had been received on the same day (from different surgeons) and were assigned sequential accession numbers.

7. CONCLUSION

Genotyping of individuals for the purpose of identity using a variety of methods is now widely accepted and can be critically applied in cases of questioned clinical specimen identification. A clinical identity issue can arise because of a potential specimen mix-up, a "floater," or even a patient in denial of his/her diagnosis. In these varying scenarios, polymorphic DNA marker analysis along with state-of-the-art technological expertise and histopathologic review can be summoned to definitively resolve identity issues in the majority of cases.

REFERENCES

1. Moenssens, A. A. *Fingerprints and the Law*. Chilton, Philadelphia, 1969.
2. Gaensslen, R. E. *Sourcebook in Forensic Serology, Immunology, and Biochemistry*. Superintendent of Documents, Washington, DC, 1983, pp. 12–14.
3. Sensabaugh, G. F. Biochemical markers of individuality, in *Forensic Science Handbook,* R. Saferstein, ed. Prentice-Hall International, Englewood Cliffs, NJ, 1982, p. 340.
4. Wyman, A. R. and White, R. A highly polymorphic locus in human DNA. Proc. *Natl. Acad. Sci. USA* **77**:6754–6758, 1980.
5. Jeffreys, A. J., Wilson, V., and Thein, S. L. Hypervariable minisatellite regions in human DNA. *Nature* **314**:67–73, 1985.
6. Jeffreys, A. J., Brookfield, J. F. Y., and Semeonoff, R. Positive identification of an immigration test-case using DNA fingerprints. *Nature* **318**:577–579, 1985.
7. Jeffreys, A. J., Wislon, V., and Thein, S. L. Individual-specific fingerprints of human DNA. *Nature* **316**:76–79, 1985.
8. Gill, P., Jeffreys, A. J., and Werrett, D. J. Forensic applications of DNA fingerprints. *Nature* **318**:577–579, 1985.
9. Kanter, E., Baird, M., Shaler, R., et al. Analysis of restriction fragment length polymorphisms in deoxyribonucleic acid (DNA) recovered from dried bloodstains. *J. Forensic Sci*. **31**:403–408, 1986.
10. Ali, S., Muller, C. R., and Epplen, J. T. DNA fingerprinting by oligonucleotide probes specific for simple repeats. *Hum. Genet.* **74**:239–243, 1986.
11. He, Z. N., Jiang, X. H., Lu, S. H., et al. A study of sex identification of trace, dried bloodstains using a Y-chromosome specific deoxyribonucleic acid (DNA) probe. *J. Forensic Sci*. **34**:346–351, 1989.
12. Honma, M., Yoshii, T., Ishiyama, I., et al. Individual identification from semen by the deoxyribonucleic acid (DNA) fingerprint technique. *J. Forensic Sci.* **34**:222–227, 1989.
13. Helmuth, R., Fildes, N., Blake, E., et al. HLA DQα allele and genotype frequencies in vario8us human populations, determined by using enzymatic amplification and oligonucleotide probes. *Am. J. Hum. Genet.* **47**:515–523, 1990.
14. Comey, C. T., Budowle, B. Validation studies on the analysis of the HLA DQα locus using the polymerase chain reaction. *J. Forensic Sci.* **36**:1633–1648, 1991.
15. Brinkmann, B., Rand, S., and Wiegand, P. Population and family data of RFLPs using selected and single- and multi-locus sytems. *Int. J. Legal Med.* **104**:8–86, 1991.
16. Jeffreys, A. J., MacLeod, A., Tamaki, K., et al. Minisatellite repeat coding as a digital approach to DNA typing. *Nature* **354**:204–209, 1991.
17. Hagelberg, E., Gray, I. C., and Jeffreys, A. J. Identification of the skeletal remains of a murder victim by DNA analysis. *Nature* **352**:427–429, 1991.
18. Gaensslen, R. E., Berka, K. M., Grosso, D. A., et al. A polymerase chain reaction (PCR) method for sex and species determination with novel controls for deoxyribonucleic acid (DNA) template length. *J. Forensic Sci.* **37**:6–20, 1992.
19. Erlich, H., Bugawan, T., Begovich, A., et al. Analysis of HLA class II polymorphism using polymerase chain reaction. *Arch. Pathol. Lab. Med.* **117**:482–485, 1993.
20. Comey, C. T., Budowle, B., Adams, D. E., et al. PCR amplification and typing of the HLA DQα gene in forensic samples. *J. Forensic Sci.* **38**:239–249, 1993.
21. Tsongalis, G. J., Coleman, W. B., Esch, G. L., et al. Identification of human DNA in complex biological samples using the Alu polymerase chain reaction. *J. Forensic Sci.* **38**:961–967, 1993.
22. Hammond, H. A., Jin, L., Zhong, Y., et al. Evaluation of 13 short tandem repeat loci for use in personal identification applications. *Am. J. Hum. Genet.* **55**:175–189, 1994.
23. Cosso, S. and Reynolds, R. Validation of the AmpliFLP D1S80 PCR amplification kit for forensic casework analysis according to TWGDAM guidelines. *J. Forensic Sci.* **40**:424–434, 1995.
24. Micka, K. A., Sprecher, C. J., Lins, A. M., et al. Validation of multiplex polymorphic STR amplification sets developed for personal identification applications. *J. Forensic Sci.* **41**:582–590, 1996.
25. Liechti-Gallati, S. and Neeser, D. Efficient and reliable PCR-based detection of the ABO blood group alleles: genotyping on stamps and other biological evidence samples. *J. Forensic Sci.* **41**:653–657, 1996.
26. Budowle, B., Smerick, J. B., Keys, K. M., et al. United States population data on the multiplex short tandem repeat loci—HUMTHO1, TPOX, and CSF1PO—and the variable number tandem repeat locus D1S80. *J. Forensic Sci.* **42**:846–849, 1997.
27. Watson, J. D. and Crick. F. H. C. Molecular structure of nucleic acids. *Nature* **171**:737–740, 1953.
28. Watson, J. D. and Crick, F. H. C. Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171**:964–967, 1953.
29. Nowak, R. Mining treasures from "junk' DNA. *Science* **263**:608–610, 1994.
30. Kan, Y. W. and Dozy, A. M. Polymorphism of DNA sequence adjacent to the human Beta-globin structural gene: relationship to sickle mutation. *Proc. Natl. Acad. Sci. USA* **75**:5631–5635, 1978.
31. Butstein, D., White, R. L., Skolnick, M., and Davis, R. W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**:314–331, 1980.
32. Sutherland, G. R. and Richards, R. I. DNA repeats: A treasuy of human variation. *N. Engl. J. Med.* **331**:191–193, 1994.
33. Tsongalis, G. J., Wu, A. H. B., Silver, H., and Ricci, A. Jr., Applications of forensic identity testing in the clinical laboratory. *Am. J. Clin. Pathol.* **112**:S93–S103, 1999.
34. Hunt, J. L., Swalsky, P., Sasatomi, E., Niehouse, L., Bakker, A., and Finkelstein, S. D. A microdissection and molecular genotyping assay to confirm the identity of tissue floaters in paraffin-embedded tissue blocks. *Arch. Pathol. Lab. Med.* **127**:213–217, 2003.
35. Southern, E. M. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**:503–517, 1975.
36. Saiki, R. K., Scharf, S., Falcona, F., et al. Enzymatic amplification of beta-globin genomic sequences and restriction analysis for diagnosis of sickle cell anemia. *Science* **230**:1350–1354, 1985.
37. Mullis, K. B. and Falcona, F. Specific synthesis of DNA in vitro via a polymerase catalyzed chain reaction. *Methods Enzymol.* **155**:335–350, 1987.
38. Templeton, N. S. The polymerase chain reaction: history, methods, and applications. *Diagn. Mol. Pathol.* **1**:58–72, 1992.
39. Sorscher, D. H. DNA amplification techniques, in *Molecular Diagnostics for the Clinical Laboratorian*, W. B. Coleman and G. J. Tsongalis, eds., Humana P, Totowa, NJ., 1997, pp. 89–101.

40. Lins, A. M., Sprecher, C. J., Puers, C., et al. Multiplex sets for the amplification of polymorphic short tandem repeat loci: Silver stain and fluorescence detection. *Biotechniques* **20**:882–889, 1996.
41. Lazaruk, K., et al. Genotyping of forensic short tandem repeat (STR) systems based on sizing precision in a capillary electrophoresis instrument. *Electrophoresis* **19**:86–93, 1998.
42. Deforce, D. L., et al. Comparison of slab gel electrophoresis and capillary electrophoresis for the detection of the fluorescently labeled polymerase chain reaction products of short tandem repeat fragments. *J. Chromatogr. A* **806**:149–155, 1998.
43. www.str.nist.gov/home.html.
44. Tsongalis, G. J. and Berman, M. M. Application of forensic identity testing in a clinical seting: Specimen identification. Diagn. Mol. Pathol. **6**:111–114, 1997.
45. Tsongalis, G. J., Anamani, D. E., and Wu, A. H. B. Identification of urine specimen donors by the PM + DQA1 amplification and typing kit. *J. Forensic Sci.* **41**:1031–1034, 1996.
46. Linfert, D. R., Wu, A. H. B., and Tsongalis, G. J. The effect of pathologic substances and adulterants on the DNA typing of urine. *J. Forensic Sci.* **43**:1041–1045, 1998.
47. Shibata, D., Kurosu, M., and Noguchi, T. T. Fixed human tissues: a resource for the identification of individuals. *J. Forensic Sci.* **36**:1204–1212, 1991.
48. Wilson, L. J., Horvat, R.T., Tilzer, L., et al. Identification of donor melanoma in a renal transplant recipient. *Diagn. Mol. Pathol.* **1**:266–271, 1992.
49. Abeln, E. C. A., Van Kemenade, F. D., Van Krieken, J. H. J. M., et al. Rapid identification of mixed up bladder biopsy specimens using polymorphic microsatellite markers. *Diagn. Mol. Pathol.* **4**:286–291, 1995.

Issues for the Clinical Molecular Pathology Laboratory

X

43 Genetic Counseling Considerations in Molecular Diagnosis

MYRA I. ROCHE

1. INTRODUCTION

The focus of this chapter is to discuss the clinical and counseling applications of molecular genetic testing. Genetic Counseling is a term most strictly applied to the communication process in which individuals and families engage with specially trained health professionals in a process of education and counseling in reference to a disability, disease, or trait that might be caused by a genetic change. This service has historically been performed in conjunction with the provision of clinical genetic services. Genetic counselors are specifically trained health professionals, certifiable by nationally recognized boards, to perform genetic counseling. Concurrent with the expansion of the molecular diagnostic laboratory capabilities to detect genetic changes, the role of genetic counselors has evolved in the more than 30 yr since its professional inception and now includes practitioners in a variety of settings, including clinical, academic, laboratory, research, industry, and public health. An extensive body of literature has accumulated relevant to the practice of genetic counseling, much of it generated by researchers in fields other than medical or clinical genetics. The literature represents a heterogeneous mixture of disciplines, philosophical and cultural assumptions, methodology, and populations sampled.

The clinical applications of genetic testing and the resulting genetic counseling considerations are presented in the context of the practice of clinical genetics services. Because valid interpretations of the literature depend on an appreciation for the differences in the education, training, and practice philosophies of those engaged in genetic counseling as well as the social and cultural mileu surrounding these services in different countries, these are described. The unique features of the genetic counseling process, misconceptions about its goals, and its historical and philosophical roots in the United States are summarized.

To understand how the provision of molecular genetic testing raises complex and difficult issues that are best addressed by the provision of genetic counseling, five clinical applications are described and analyzed. The uses of genetic testing, their potential implications for relatives, and determinations of appropriate candidates, especially when children are considered, are discussed. Asymptomatic children are controversial candidates for diagnostic, carrier, predictive, and susceptibility genetic testing. Determining which indications are appropriate for prenatal diagnosis is becoming increasingly difficult, as the ability to detect genetic changes expands. The roles of the laboratory, the ordering physician, and the patient and family are examined to identify barriers to the accurate interpretation, communication, and the dissemination of genetic test results. These barriers include: the complexity of the human genome, the level of knowledge of both provider and patients, about genetic principles and attitudes and previous experiences with genetic disorders and genetic counseling. Factors that correlate with patient interest in and use of genetic testing and the potential psychological impacts of testing are discussed. It is expected that these issues will become increasingly more complex as genetic testing is marketed directly to families.

2. INTRODUCTION TO CLINICAL GENETIC SERVICES

2.1. CLINICAL GENETICS AND GENETIC COUNSELING

Donnai *(1)* cites the World Health Organization's broad definition of genetic services as ". . . health measures implemented to help people with a genetic disadvantage and their families live and reproduce as normally as possible . . ." This has historically been accomplished by the intertwined practices of clinical genetics and genetic counseling. Clinical genetics is the branch of medicine that integrates various diagnostic tools and their results into clinically meaningful answers in the effort to diagnose symptomatic individuals with genetic conditions and make appropriate plans for their optimal care and management. In addition to either providing or arranging for appropriate medical management, the educational and psychological needs of the patient and family are addressed. The delineation of the clinical features of a genetic condition, in conjunction with information provided by a comprehensive family history and other diagnostic testing, provides useful guidance in determining which molecular diagnostic tests are likely to yield definitive answers and which members of a family are appropriate testing candidates. Genetic counseling is the process by which the medical and genetic information about the disorder and

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

their implications are translated and communicated to the individual and/or family in a manner that optimizes their understanding and aids their subsequent psychological adjustment.

Different models of providing genetic services, the way health care is subsidized in a society, and the different training, experience, and demographic factors such as age, gender, and cultural background of the practitioners influence the delivery of genetic services. These, in turn, can modify patient and societal views on the acceptable use and dissemination of genetic information. An example of this relationship is the provision of genetic services through genetic registries as in the United Kingdom with patient acceptance of the sharing of genetic information outside the traditional provider–patient relationship is *(2,3)*.

2.2. CANDIDATES FOR GENETIC COUNSELING A list of appropriate candidates for genetic services has been described and is available through the Internet (www.geneclincs.org) as well as in the published literature *(4,5)*. Some candidates for genetic counseling are individuals affected with a genetic (or possibly genetic) condition who seek the diagnostic capabilities of the clinical geneticist who may order genetic testing. Others are unaffected relatives who request information about their genetic status with regard to an established genetic diagnosis in a family member. In these latter cases, the educational and psychological components become the focus. Genetic counseling is also indicated for unaffected individuals and their families who meet various criteria for genetic screening (i.e., maternal age, ethnicity) or to those families in which a member has a chronic or common disease, like cancer, that could be the result of a genetic predisposition.

2.3. GENETIC COUNSELING AND MOLECULAR DIAGNOSIS The genetic counseling considerations stemming from molecular diagnosis of genetic disorders grow increasingly broader and more complicated in response to the expanded ability to detect a greater number and variety of alleles that deviate from normal or wild type. Genetic disorders, those caused, at least in part, by a genetic change, encompass a wide spectrum of phenotypes. These, in turn, are determined by a variety of genotypes. Although the discussion in this chapter will focus on issues arising from genetic changes detectable by molecular techniques, it should be recognized that this restriction is an artificial one. Definitive diagnostic and/or carrier status of many genetic disorders can often be determined by one or more of several different modalities, including biochemical assays, cytogenetic analysis, physical examination, imaging, and pedigree analysis. Because of the varying potential of any of these to reveal the genetic status of an individual, their effective and beneficial use demands the user be familiar with the possible implications of the knowledge obtained, be trained in their appropriate interpretation and application, be able to communicate the findings or results in a comprehensible way to families, and be competent to address the accompanying psychological components that inevitably arise. A recent controversy surrounding the privacy rights of relatives who are included when a family history is obtained in research projects (American Society of Human Genetics: Family pedigree info and privacy; Should family members about whom you collect only medical history information for your research be considered "human subjects"? http://www.ashg.org/genetics/ashg/pubs/policy/pol-38.htm) and in publications *(6,7)* has heightened awareness of this potentially sensitive information.

2.4. GENETIC TESTING AIDS IN THE DIAGNOSIS, COUNSELING, AND MANAGEMENT OF FAMILIES WITH GENETIC CONDITIONS BUT RAISES ETHICAL DILEMMAS The ability to correctly diagnose some genetic disorders preceded the development of molecular techniques, but the increased use of an expanding array of molecular testing has greatly enlarged the diagnostic repertoire of clinicians. Genetic testing has the potential to provide information that is simply not attainable by previous techniques. In some cases, the acquisition of this information precedes our ability to interpret it in a clinically meaningful way. In other cases, it provides answers to questions that were not explicitly asked and about people for whom this information was not requested. The use and dissemination of this comprehensive information has ethical implications. Many genetic counseling dilemmas arise from conflicts between the rights and responsibilities of those who obtain genetic testing results and their relatives. The limitations and conflicts discussed in many articles *(8–13)*, although a concern, should not obscure the tremendous diagnostic power that molecular genetic techniques possess. Their appropriate application has spared many families from costly and invasive procedures, lengthy delays in achieving a diagnosis, and resolution of some of the uncertainties regarding the health and well-being of themselves or their relatives *(14)*.

2.5. EFFECTIVELY COMMUNICATING A GENETIC DIAGNOSIS INVOLVES BOTH CONTENT AND MANNER The manner in which information about a genetic diagnosis in a family is delivered is not forgotten. When done inappropriately, it is not forgiven easily *(15)*.As Cary *(16)* phrases it,

> . . .one of the most common and challenging dilemmas facing practitioners of genetic medicine is the provision of accurate, realistic, and balanced information to families at the times of initial diagnosis. It is axiomatic that the information needs to be . . . sensitive to the person's culture, and is understandable to the family. Characterization of the course of the disorder requires that the practitioner be both realistic (discuss known outcomes and potentially serious complications) and optimistic (promoting hope and a positive outlook)…while presenting an overall picture that avoids stigmatization. . .

As an example, parents seen in a genetics clinic in the United Kingdom described the use of specific words as problematic; the word "syndrome" was frightening and the use of eponyms was confusing *(17)*.

In interviews with parents of children born with short stature, the manner in which the diagnosis was conveyed had a major impact on parental adjustment and acceptance that was "vividly remembered" and colored their future relationships with health professionals *(18)*. For the majority of those parents, the diagnosis was conveyed by a physician with inadequate knowledge about the disorder and its prognosis, by the use of stigmatizing and unfamiliar terms, and by accompanying these with distressing, clinically oriented pictures. Positive information such the

expected normal life span and normal intelligence and the chance nature of new mutations (which can help alleviate parental guilt) were rarely described. The need for accurate written information about the condition, its natural history, and the provision of sources of social support and services was noted.

2.6. CLASSIFICATIONS OF GENETIC TESTING INDICATIONS Genetic testing can be done for many purposes, each of which has its own array of implications. Classifications like the one following have been proposed and used by many authors *(19–21)*. Readers with a limited background in medical genetic terms and practice might benefit from a comprehensive textbook for an explanation of commonly used terms *(4,5)*. Although this classification emphasizes the detection of abnormal alleles, the opposite result is also possible but sometimes overlooked (i.e., the detection of normal alleles). The interpretation of this result will vary depending on whether or not a causative abnormal allele has already been identified in the family, whether the absence of a relatively common mutation will allow revision of an *a priori* risk, or whether the phenotype is associated with limited allelic or locus heterogeneity. Genetic testing can be grouped into the following categories: (1) diagnostic testing, (2) carrier testing, (3) predictive testing, (4) susceptibility testing, or (5) prenatal testing.

Diagnostic testing detects one or more specific altered allele(s) in a symptomatic individual and confirms a clinically suspected disorder or disability. Examples of diagnostic testing include GJB2-related sensorineural hearing loss, mental retardation resulting from Fragile X syndrome, and choreoathetoid movements caused by Huntington's disease.

Carrier testing detects altered alleles that confer an increased risk of transmission to future generations, in whom it has the potential to cause disease/disability. This term is conventionally reserved for asymptomatic heterozygotes for conditions or traits inherited in autosomal recessive or X-linked patterns. Examples of carrier testing include identification of heterozygotes for cystic fibrosis and Fragile X syndrome. Although the inheritance of Fragile X syndrome is complex, with varying degrees of symptoms in heterozygous females depending on the size of the repeat expansion, it is still customary to refer to women with a premutation as carriers.

Predictive testing detects altered alleles for highly penetrant disorders in individuals who are not currently clinically affected but are expected to show some symptoms at some unspecified time in the future. Examples of predictive testing include Huntington's disease and familial adenomatous polyposis.

Susceptibility testing detects an altered allele that substantially increases the chance that an individual who is not currently clinically affected will show symptoms at some unspecified time in the future. In contrast to the predictive testing, the disorders in this classification show incomplete penetrance resulting in a decreased proportion of those with the altered allele exhibiting the phenotype. An example of susceptibility testing would be identification of mutations in BRCA1.

Prenatal testing detects altered allele(s) of any of the above categories of testing performed during pregnancy. It is most often performed because of an increased risk of an abnormality either because there is a family history of a genetic condition, a screening criterion is met (such as increased maternal age or maternal serum screening), or when an abnormal phenotype is present. Newer techniques such as preimplantation genetic diagnosis (PGD) can be considered under this category.

3. USES OF GENETIC TESTING

3.1. GENETIC TESTING CAN ESTABLISH OR CONFIRM A DIAGNOSIS The following example shows the extent to which various diagnostic modalities play a greater or lesser role in the accurate diagnosis of a genetic disorder. A greatly elevated creatine kinase in a 4-yr-old boy who has proximal muscle weakness and a maternal uncle with Duchenne muscular dystrophy (DMD) are, taken together, highly sensitive and specific indicators of the boy's probable diagnosis. Although one can be reasonably confident of this diagnosis in the absence of genetic testing, the ability to detect a specific genetic change in this boy (referred to as the proband) has multiple advantages. When the result is highly predictive of disease, genetic testing acts as confirmatory evidence for the diagnosis and serves as a gold standard. This ability varies with different disorders and with factors such as penetrance. In the absence of clearly indicative clinical signs and for cases in which the proband is the only affected individual in the family, genetic testing might be the only way a diagnosis can be established. Successful identification of the molecular defect often obviates the need for additional diagnostic testing that can be invasive, painful, distressing, and costly *(22,23)*. The likelihood that a specific genetic change will be identified should be discussed with a family before embarking on diagnostic testing so that their expectations about whether the testing will confirm a suspected diagnosis are realistic. Complexities of the genome that decrease this likehood such as locus and allelic heterogeneity are poorly recognized and poorly understood concepts in the general population.

3.2. GENETIC TESTING CAN HELP DEFINE THE NATURAL HISTORY OF A DISORDER Depending on the result and the disorder in question, defining the molecular defect might help, to a degree, to predict the progression of the disease. This is especially true in disorders with a close genotypic–phenotypic correlation and in the absence of other affected relatives. When several relatives are affected, knowledge of their clinical histories can shape these phenotypic predictions as well. For an increasing number of genetic disorders, the molecular definition helps to direct treatment, educational interventions, or other managements. This potential underlies the expectation that genetic testing will be used in the detection of susceptibilities to common diseases and play an increasing role in the prevention or modification of subsequent sequelae *(20,24–27),* although some have questioned the rate and degree to which genetic testing will be integrated into primary care practices *(28)*.

3.3. GENETIC TESTING CAN PROVIDE INFORMATION FOR RELATIVES In addition to the potential diagnostic prowess of genetic testing, the unique power of the ability to detect a specific genetic change in the proband (and, in this case, his uncle) lies in the ability to apply this knowledge to a broader group of individuals—the unaffected relatives. Genetic testing has the potential to precisely reveal the genetic status of family members with respect to a particular gene. In this example, the

detection of a deletion in the proband allows the carrier status of his sisters and maternally related relatives to be determined. The revelation to asymptomatic individuals that they have a genetic change with implications for very personal and sensitive issues, such as conceiving children with significant problems, requires empathetic counseling and a keen awareness of typical mistaken assumptions about the prevalence of genetic changes in the population. The detection of a causative genetic change also makes prenatal diagnosis technically possible, although controversial for some phenotypes. When multiple relatives are tested, it is likely that there will be discrepant results that, especially among close relatives such as siblings, can cause psychological distress *(29,30)*. This possibility should be anticipated and discussed during the counseling session.

3.4. CURRENT CONTROVERSIES: TESTING IN CHILDREN AND CONTACTING RELATIVES One controversy stemming from the familial nature of genetic changes involves whether or not asymptomatic children or adolescents should have predictive *(24,31–36)* or carrier testing *(37–43)*. Although restrictive and cautionary opinions and statements were initially issued *(31,44,45)*, there is currently a trend for greater parental involvement in the decisions and increased flexibility with regard to older adolescents *(46,47)*. A second topic that has generated much discussion is the extent to which relatives are informed of a genetic test results that has implications for them *(39,48–53)*, how genetic professionals can facilitate the dissemination of genetic information in a family *(54–57)*, and whether they have a duty to disclose genetic information to relatives of the proband without consent *(2,3,58–65)*. These issues will be discussed in more detail as they relate to the clinical cases presented.

3.5. THE BENEFITS OF DIAGNOSTIC TESTING The limitations and potential for misinterpretations of genetic testing results by both health care professionals *(66–72)* and patients *(33,73–77)* are topics well represented in the genetic counseling literature. Because examples of predictive, carrier, and susceptibility testing of asymptomatic relatives are overrepresented in the literature, the diagnostic power of genetic testing might be obscured.

Genetic testing has the unsurpassed potential to determine a diagnosis where none could be made in the past. Although families report mixed feelings after a genetic diagnosis is made (such as relief that an answer is obtained yet sadness because it implies fewer treatment options and a sense of permanence), many families continue to search for a diagnosis and experience frustration at the barriers preventing early diagnosis *(78–80)*. A diagnosis provides information to the patient on etiology, prognosis, potential for treatment, recurrence risks, life expectancy, and anticipatory guidance. In addition, a diagnosis provides support for the patient in the form of acceptance of the individual by others in the family and society, absolution of parental responsibility, as well as increased ability to obtain needed services *(79,81)*. Optimally, the knowledge of all of these components, the clinical and biochemical evidence, the family history, and the molecular laboratory results, combine synergistically in the attempt to establish an accurate diagnosis and provide a basis for genetic counseling of the family. Genetic testing results, when unaccompanied by these other critical components, including genetic counseling, remains powerful but has a greatly magnified potential for misinterpretation, misapplication, and misuse regardless of whether the individual is symptomatic or asymptomatic.

4. COMPONENTS OF THE GENETICS COUNSELING PROCESS

Comprehensive texts about the practice of genetic counseling are available both for the broad range of genetic counseling indications *(82)* and specifically for predisposition to cancer *(83)*. Specific recommendations for genetic evaluations and counseling have been published *(84,85)* and national standardized guidelines are being developed *(86)*. The elements and techniques in taking a genetic family history are detailed by Bennett *(87)*, including an exhaustive compendium of standardized symbols and their appropriate use and examples of interpretations. The psychosocial elements and historical roots of genetic counseling are thoroughly discussed by Kessler *(88)* and Weil *(89)*.

The components of a genetic evaluation and counseling session can differ tremendously depending on the specific indications. A list of these components can be found on the Internet (www.geneclinics.org) as well as in the literature *(82,83,85)*. The challenge of providing competent genetic counseling stems from the unique combination of heterogeneous components that each case presents. These include genetic factors (locus and allelic heterogeneity), phenotypic variability, and differences in family structure, beliefs, and functioning. Even in cases in which the genetic and phenotypic heterogeneity are limited, the diagnosis of a genetic condition or disability in an individual or a relative has multiple effects on a family. Among affected families some of these effects are common while some will be unique.

The specific components of the genetic counseling process include (1) assessing the purpose of the visit and the family's needs, assumptions, and expectations, (2) gathering information about the medical, family, psychosocial, and other relevant history, (3) collecting, analyzing, and synthesizing relevant information, including physical signs, (4) communicating the diagnostic considerations and the possible modes of inheritance, as well as other information about appropriate genetic testing and/or other medical testing, (5) providing psychological support and assessing whether additional support might be indicated, (6) recommending appropriate testing and follow–up for medical, genetic, educational, and psychological needs, and (7) providing written documentation of the visit. Each component of the genetic counseling process can include other relevant topics, such as discussion of future reproductive options, strategies to inform other relatives about their potential risk, the potential for discrimination and stigmatization of the affected individual, and the opportunity to participate in research studies or clinical trials.

4.1. DIAGNOSTIC GENETIC TESTING The use of genetic testing to confirm a diagnosis that is suspected based on clinical features is, perhaps, the most easily understood and least controversial. This testing can be done either in the presence or the absence of a positive family history. There is a continuum in how strongly or weakly a specific diagnosis is suspected. In some cases, the clinical features of the disease readily lend themselves to the diagnosis. In others, the phenotypic range might be broad and nonspecific and/or might be sufficiently

rare such that a range of diagnoses is considered. Diagnostic genetic testing for cystic fibrosis and Fragile X syndrome represent different ends of this spectrum. The diagnosis of cystic fibrosis is usually highly suspected based on clinical features, and abnormal sweat chloride levels are often determined before genetic testing is undertaken. In contrast, Fragile X testing might be done in conjunction with a number of other tests (such as cytogenetic and biochemical tests) to investigate the presence of the nonspecific phenotype of mental retardation. The use of GJB2 testing for sensorineural hearing loss is becoming more common in cases of severe to profound, congenital loss. Because there are usually no other health problems associated with this phenotype, it does not fit the usual profile of a genetic disease. Thus, prenatal testing for this condition is controversial. As genetic testing becomes more widely employed, the phenotypic spectrum of many disorders, once believed to be relatively homogenous, has greatly expanded. An example is the recognition that two altered alleles of the CFTR gene can result in normal sweat electrolytes and the congenital bilateral absence of the vas deferens (CBAVD) with absence pulmonary involvement as the one of clinical finding.

4.2. FAMILY HISTORY COLLECTION One crucial but often neglected aspect of the clinical evaluation involves comprehensive analysis of the family history *(87)*. The presence of other affected relatives, in conjunction with features detected by clinical examination or other testing, facilitates the selective targeting of genetic testing. Several studies have shown that families are not offered appropriate genetic testing because of the failure of their physician to take a comprehensive family history *(90,91)*. Experience with a genetic disorder in the family can temper reactions to the genetic testing results because of their awareness, at least on some level, that the elicitation has an increased chance of the disease. The differing reactions between those familiar with the condition and those with no experience *(92,93)* have important implications for genetic counseling in the context of population screening. It should be recognized that relatives vary in their knowledge of the presence of a genetic disorder in the family because of different disclosure rates and patterns of selectively informing specific relatives. *(2,50,53,55,56,94–98)*. In individual's relatives for many genetic conditions, especially those inherited in an autosomal recessive pattern and in families with few children, there are no other affected relatives. The abstract explanation of how a genetic condition can be inherited in the absence of a family history can be a difficult concept for families to understand. Many families have developed their own theories for the causation of a disorder in the family and this can be culturally determined *(17,99,100)*. If these interpretations are not explored, the scientific explanations might well be discounted *(101,102),* although, in some individuals, both their personal explanations and the accepted scientific explanations seem to be able to coexist *(101)*.

5. DEFINITIONS, GOALS, AND PRACTITIONERS OF GENETIC COUNSELING

5.1. THE TERM "GENETIC COUNSELING" The term "genetic counseling" is confusing to both health professionals and the public *(103)*. In part, this is because of the juxtaposition of two terms—one scientific and one psychological—that, to the uninitiated, seems paradoxical. This juxtaposition was intentional. When Sheldon Reed first proposed the term in 1947 *(104)*, he expected that genetic counseling would allow the growing knowledge of human genetics to be used to "… help individual families…," but did not want it to "…become the tool of any government population program …." His book, *Counseling in Medical Genetics (105),* was written for primary care physicians, as he believed they would be providing most of the genetic counseling *(106)*. This prediction appears to have been correct. Two recent studies found that fewer than 20% of families who meet standard criteria for genetic counseling were referred and over half of the parents receiving information about the recurrence risks following the birth of an affected child were informed by their general practitioner *(103,107)*. These are very likely to be underestimates. Fifty years after Reed's prediction the literature is filled with admonitions about the need to educate primary care physicians in anticipation of expanding genetic testing options *(108–111)*.

5.2. EXPANDED DEFINITIONS OF GENETIC COUNSELING Until the 1960s, many genetic disorders could not be reliably diagnosed. The recognition of various chromosomal aneuploidies, the biochemical testing for disorders such as Tay–Sachs disease, and the invention of prenatal diagnostic techniques gradually evolved in the 1960s and 1970s concurrent with the establishment of many genetic counseling services. In 1975, the American Society of Human Genetics proposed an expanded definition of genetic counseling to specifically include: both educational and psychological aspects; a description of genetic counseling as a communication process rather than one with a preventative goal; an emphasis on the understanding of medical and genetic information by families; a need for the discussion of options; the pivotal role of decision-making; and the need for enhancing psychological adjustment *(112)*. Biesecker and Marteau *(113)* described the genetic counseling process as one that "may seek to facilitate informed and autonomous decision-making, appreciation of the inheritance of a genetic condition, integration of genetic information into a useful framework, or improvement in the emotional well-being of those affected or their families."

5.3. ASSUMPTIONS UNDERLYING GENETIC COUNSELING Pagon *(21)* defined genetic counseling as a process that "focuses on the" assessment of risk, education for the at-risk person and family about the disorder and its management, education about reproductive options and, psychological, and emotional support for largely untreatable diseases. . . " Noting that the philosophical tenets underlying genetic counseling were ". . . rooted in the social sciences. . .," Pagon stated that genetic counseling for disorders of ". . . high medical impact. . ." is ". . . not intended to modify behavior. . .," but rather ". . . to enhance informed decision making. . ." and ". . . to promote timely and appropriate medical intervention, when available. . ." *(21)*. Pagon added that the use of nondirective techniques "facilitates personal decision making when only supportive treatments (for disorders) are available" *(21)*. Walker *(114)* listed the major assumptions underlying the provision of genetic counseling. These included (1) the voluntary nature of genetic services, (2) promotion of equal access, (3) disclosure of all relevant information, (4) confidentiality and protection of privacy,

(5) integration of the psychological component, and (6) nondirectiveness. These are also enumerated in the Code of Ethics of the National Society of Genetic Counselors *(115)* and provide the foundation for the ethos of genetic counseling.

5.4. THE GOALS OF GENETIC COUNSELING The philosophical basis underlying genetic counseling encompasses the unorthodox mixture of the two major goals: (1) provision of information that has both medical and personal relevance and (2) the responsibility of evoking and responding to the psychological needs of the client that result from this provision. A hybrid arising from quite disparate parentage, genetic counseling straddles, on one hand, a directive, expert-driven model commonly employed in traditional medical settings, whereas, on the other hand, espouses a Rogerian client-centered approach characteristic of psychotherapeutic approaches *(116,117)*. The differences in social and cultural attitudes prevalent in the United States in contrast to many European countries during the early development of the field also played a role *(118,119)*. These distinctions are revealed by differential responses of genetic professionals world-wide to ethical challenges that occur in the provision of genetic services and counseling *(120,121)*.

5.5. DUAL ROLES LEAD TO ETHICAL CONFLICTS Conflicts between the two underlying models characterize many of the ethical conflicts inherent in the provision of genetic services. The decision to utilize genetic services, especially genetic testing, should be voluntary, yet newborn screening of an increasing number of genetic disorders is mandated. All relevant information about genetic testing and genetic disorders should be disclosed, but the revelation of unanticipated false paternity might not be in the patient's best interests. Privacy and confidentiality should be protected, but genetic information can have profound implications for relatives who might be unaware of the existence of relevant genetic knowledge gained by genetic testing of an individual. McCarthy-Veach et al. *(122)* summarized recurring professional and ethical issues identified by a focus group of experienced genetic counselors. These issues included (1) limited usefulness of testing information, (2) equitable resource allocation, (3) value conflicts, (4) nondirectiveness, (5) professional identity, (6) issues regarding diversity, (7) emotional reactions to clients, (8) discrimination, and (9) how to deal with colleague errors *(122)*.

5.6. PROVIDERS OF GENETIC COUNSELING VARY IN THEIR TRAINING The training and expertise of providers of the various components of genetic counseling services differ dramatically between countries, resulting in distinct differences in goals and methods. The most common provision model in the United States and Canada is a team including genetic counselors (trained at the master's level) and clinical geneticists (physicians with specialty training in genetics). Other health professionals might be involved, most notably nurse specialists *(123),* who are more prominent in the United Kingdom where genetic counseling training programs have more recently begun. Medical geneticists, some of whom also provide genetic counseling, and laboratory-based specialists such as cytogeneticists and molecular and biochemical geneticists, are integral to the comprehensive delivery of genetic services.

5.7. GENETIC COUNSELOR TRAINING IN NORTH AMERICA In the United States and Canada, genetic counselors are master's-level health professionals who are trained in specifically designed programs meeting requirements set by the American Board of Genetic Counseling (ABGC) or the Canadian Association of Genetic Counsellors (CAGC) for educational and clinical expertise. Graduating students are expected to demonstrate expertise in the multiple components that constitute practice-based competencies *(124)*. During the past 10 yr, eligibility requirements for training program accreditation and board certification of genetic counselors have become increasingly standardized and more restrictive, and recertification is now mandatory for genetic counselors certified in 1996 or later. Demographically, genetic counselors are primarily white (approx 94%) and female (approx 96%) although efforts continue to diversify the profession (Professional Status Survey of the National Society of Genetic Counselors, 2000, www.nsgc.org).

5.8. CLINICAL GENETICISTS AND OTHER MEDICAL GENETICISTS Clinical geneticists who are physicians with primary training in (most commonly) pediatrics, obstetrics, or, more recently, internal medicine also provide genetic counseling. These clinicians have advanced training in the diagnosis, management, and care of individuals with genetic conditions. A clinical geneticist often has a specific area of expertise. Specialties in clinical genetics include emphasis in dysmorphology, metabolic disorders, neurogenetics, prenatal diagnosis, or cancer. Clinical geneticists, along with other medical geneticists, are certified by the American Board of Medical Genetics.

5.9. THE CURRENT SUPPLY AND PROJECTED NEED FOR GENETIC COUNSELORS Currently, there are approx 1800 board-certified genetic counselors (http://www.abgc.net/). In 2002, it was estimated that there was 1 genetic counselor per 200,000 people in the United States. Although there are fewer than 1 counselor in 1,000,000 people in Europe, many are not specifically trained, and some countries have no genetic counselors *(125)*. It is generally accepted that increases in genetic testing capabilities will result in an increased demand in their routine use in the clinical work-up of patients, such that the number of genetic counselors and clinical geneticists with appropriate training and certification will become increasingly inadequate. Tremendous effort will be required to educate the necessary numbers of primary care physicians, nurses, allied health, and public health practitioners in the nuances of taking a family history and in basic genetic principles *(126–131)*. To some extent, these efforts have begun *(132)*. Nonetheless, some have wondered why little funding exists for a concurrent effort to train more geneticists and genetic counselors *(133)*. With ongoing training efforts, the number of newly trained genetic counselors continues to rise, at a steady, although modest rate (250 were certified in 1996; over 400 in 2002). The number of newly trained clinical geneticists has plateaued, and since 1999, the number has decreased (from about 125 in 1996 to 69 in 2003, for a total of slightly over 1000). This decline is mirrored in the other specialties of medical genetics *(133*; http://www.faseb.org/genetics/abmg/abmgmenu.htm).

5.10. FUTURE TRENDS IN GENETIC SERVICE DELIVERY As diagnostic capabilities have increased, genetic testing has become available for disorders typically managed in a wide range of specialty clinics—most notably, cancer. Physicians who have

primary training in areas other than genetics have formed teams with genetic service providers to provide comprehensive services for patients and families. As members of such teams, genetic counselors provide the expertise in genetic principles and psychological counseling—a model that has dominated the delivery of prenatal diagnosis for some time. Although it is expected that the continued expansion of genetic testing will create additional markets for genetic counselors, issues such as reimbursement will affect the services delivered and the time allotted for each. Currently, in all but one state, genetic counselors are not licensed and cannot bill independently for their services. This is expected to change in the near future and could fundamentally change the settings in which genetic counseling is delivered.

5.11. PHILOSOPHICAL AND SOCIETAL DIFFERENCES AMONG PROVIDERS OF GENETIC COUNSELING As the preceding discussion indicates, the term "genetic counselor" is used in a nonspecific way, causing confusion as to which kind of practitioner is actually meant when the term is used in the genetic counseling literature. Genetic counselors and clinical geneticists in North America share the same basic philosophical ethos. Still, Pencarinha et al. *(134)* found some evidence to suggest that genetic counselors more strongly support patient autonomy than medical geneticists. The differences between these two groups are probably modest compared to those between genetic professionals and nongeneticist health care providers and between geneticists in various countries *(113,120,135)*. The structure, delivery, and reimbursement of genetic services in different nations also vary and affect interpretation of outcome studies of genetic counseling.

5.12. NONDIRECTIVENESS The use of the term "nondirectiveness" is often used as a proxy for the manner in which genetic information is ideally delivered concerning decisions made on the basis of personal values rather than on scientific or medical data. Typically but not exclusively, these involve reproductive issues. Nondirectiveness as it relates to the practice of genetic counseling has been the subject of a large number of discussions that seek to explain its intent *(122,136–138)*, question its feasibility *(139)*, disprove that it occurs in practice *(140,141)*, vilify its attempted use *(142)*, or propose that it be abandoned *(143)*. Kessler *(88)* defined it not so much as a lack of influence but rather as a lack of the "intent to direct decisions in a specific way." Most genetic counselors and geneticist believe that it is important and agree that it is the appropriate approach in genetic counseling *(146,149)*.

5.13. EXAMPLES OF WAYS IN WHICH GENETIC COUNSELING DEVIATES FROM NONDIRECTIVENES Examples in which parts of the genetic counseling process deviate from a nondirective approach are highly visible and widely accepted. These include recommendations made for medical surveillance, specific warnings against pregnancy exposures, and advising the use of folic acid supplementation before pregnancy. The mandatory nature of newborn screening and the lack of informed consent can also be viewed as directive *(113)*. However, this omission is becoming increasingly questioned *(145)*.

5.14. SPECIFIC OUTCOMES OF GENETIC COUNSELING ARE DIFFICULT TO DEFINE AND EVALUATE Early studies of genetic counseling were narrowly defined as an educational process whose outcomes could be measured as the amount of knowledge learned and the subsequent "informed" reproductive decisions made *(112,146)*. The misinterpretation persists that the goals of genetic counseling are limited to (1) providing information about disease etiology and (2) estimating risks for recurrence. Although genetic counseling is increasingly defined in broader terms, the effectiveness of the educational function is an important one and continues to be the subject of evaluative studies. Traditional counseling methods *(147)* as well as other educational modalities such as brochures *(148)*, videotapes *(149)*, audiotapes *(150)*, and interactive computer programs *(151,152)* have also been evaluated for their educational effectiveness.

5.15. THE RATE OF INCREASED KNOWLEDGE OF RISKS ON REPRODUCTIVE DECISIONS Several early papers examined the relationship between knowledge gained through genetic counseling and subsequent reproductive decisions made by patients, with the assumption that clients who understood their risks would act accordingly in family planning decisions. However, the decisions made by couples after genetic counseling varied. Several factors were identified as contributory, including parental perception of the burden of the disorder *(146,153–155)*, previous experience with a specific disorder *(156,157)*, the birth order of the affected child *(158,159)*, and the availability of prenatal diagnosis *(160,161)*. Some believed that risk information is interpreted rather than recalled as discrete numbers *(159,161–164)* with varying impact on decision-making. Finally, some families chose to have additional children following genetic counseling because they were reassured by the results *(165)*.

5.16. MEASURING THE WRONG OUTCOME Defining the expected outcome of genetic counseling by measuring the reduction of the prevalence of the disorder contradicted the espoused nondirective philosophy *(141)* and revealed the inherent inconsistencies between goals of prevention and those of improved psychological well-being. Specific reproductive decisions would affect the former goal but not the latter as long as the decision was "… congruent with the individual's values and beliefs …" *(137)*. The inability of genetic counseling to have much effect on couples' subsequent reproductive behavior has provided evidence that genetic counseling is ineffective, a view that some still hold today *(166)*. The complexity of the process of genetic counseling might mean that even a well-designed quantitative instrument is insufficient to provide an accurate measure, and qualitative methods are becoming more common *(167,168)*.

5.17. RECURRENCE RISKS COMMUNICATED TO FAMILIES ARE VARIABLE AND COMPLEX A family's recall of the recurrence risk tends to be highly dependent on the preciseness with which this risk measurement was delivered; qualitative recall (high, moderate, low) was usually superior to recall of exact numbers. The assumption that most recurrence risks are single, discrete descriptors or number values is naive. Risk estimates for a genetic condition can vary from an absolute quantity based on Mendelian principles (i.e., 50%) to a range of numbers used when the etiology of the disorder cannot be specifically defined. These latter phenotypes have heterogeneous etiologies requiring the use of empiric recurrence rates that are population

based. The ability to predict the occurrence of a specific phenotype (including severity and age of onset) is unreliable, yet this is equally important to families. The ability of physicians and families to understand probabilistic information is probably routinely overestimated *(169)*.

5.18. MEASURES OF OUTCOME OR PROCESS? In light of the difficulties in measuring genetic counseling effectiveness, Clark et al. *(170)* and Shiloh et al. *(171)* have suggested that the process rather than the outcome be evaluated. More recent studies have focused on client satisfaction *(172)* and how the process assists with decision-making and aids psychological adjustments *(122,173)*. When counseling is provided for conditions for which treatment is available, outcome measures such as changes in behavior have been used *(174,175)*. Biesecker and Marteau *(113)* discussed the problem of performing valid outcome studies when there is a " lack of consensus as to the goals" of genetic counseling, the variability between settings and indications (prenatal vs postnatal; treatable vs not treatable), and differences in the time allocated to education vs psychosocial support during a session. The relative lack of studies in which the session was observed, documented, and subjected to content analysis has contributed to a lack of information about the process, and reliance on client or providers' interpretations has been the rule until recently *(172,176,177)*.

5.19. GOALS AND EXPECTATIONS MAY DIFFER BETWEEN PROVIDER AND CLIENT Although providers are fairly consistent in their goals for a genetic counseling session *(178),* patients (and their referring physicians) have varied expectations *(122,164,172,176,179)*. Some clients expect information, explanations, advice, reassurance, and help in making decisions *(173)*. Others do not know what to expect *(176–178)* and are unprepared for the amount of information typically required from them. Some do not recognize that they have had genetic counseling, as it often occurs in the context of a medical evaluation. This uncertainty conflicts with genetic counselors' expectations that the client will actively participate in the session. A comparison of a client visit to a genetic counselor to that of a patient visit with a primary care physician revealed several differences, including the amount of time spent with the client, the manner and thoroughness of the information presented, and the ability to speak in layman's terms *(178)*. The provision of written information before *(177)* and after *(179,180)* a genetic counseling visit is customary *(82,181)*. These differences hold important implications for patient care, if as predicted *(109,111)* primary care physicians become the main providers of genetic testing interpretation and counseling.

6. GENETIC COUNSELING AND GENETIC TESTING

In many cases, accurate and clinically valid interpretations of genetic testing results are complicated. This stems in part from the complexity of the underlying structure of the genome and the limits in predicting a phenotype from a genotype. These unique characteristics differentiate genetic testing from other clinical testing. Further complications relate to an insufficient knowledge and appreciation of the complexity of genetic principles by many ordering physicians, other health care providers, and patients. Interpretations are also colored by limited previous experiences with genetic disorders and assumptions and expectations about the deterministic quality of genetic changes.

6.1. GENETIC TESTING AND THE NEED FOR GENETIC COUNSELING Recommendations citing the need for genetic counseling to be performed in conjunction with genetic testing have become standard *(21,84,182–185)*. The need for genetic counseling parallels the necessity of the oversight for genetic testing in patients (Enhancing the oversight of genetic tests: Recommendations of the Secretary's Advisory Committee on Genetic Testing, July, 2000, http://www4.od.nih.gov/oba/sacgt/gtdocuments.html). Green and Botkin *(186)* recommended that all tests with potential to cause "stigmatization, family discord, and psychological distress" should be interpreted by special trained health providers and undertaken only with the "highest standards of informed consent and privacy protection." It is difficult to estimate to what extent testing and counseling remain uncoupled, although a current survey could provide these data (CDC National Society of Genetic Counselors, www.nsgc.org). Biesecker and Marteau *(113)* observed that, in the past, commercial laboratories expected that patient samples would be obtained only after genetic counseling had occurred. Current laboratory policies of accepting samples from a variety of sources and the commencement of direct to consumer marketing "…shifts the burden of ensuring appropriate pre and post test education and counseling be done to the ordering physician…" *(113)*. Documentation of how often this is provided and to what degree it is done effectively has not been reported. Recommended guidelines for obtaining informed consent prior to obtaining a sample for predictive or susceptibility testing have been published *(149)*. However, McGovern et al. *(187)* reported that nearly 25% of samples obtained by genetic professionals themselves were collected without informed consent. A survey of laboratory directors reveals skepticism regarding the ability of ordering physicians to adequately interpret testing results *(188)*. Inadequate knowledge of genetic principles might be at least partially responsible, although this is improving and is correlated with physician specialty *(69)*. The limited accuracy and validity of some of the tests and the "…need to provide informed consent and genetic counseling for both those being tested and their relatives, whose genetic status might be inadvertently revealed by testing…" were cited as concerns *(188)*.

6.2. GENETIC HETEROGENEITY AND MISINTERPRETATIONS OF GENETIC TESTS Citing the probabilistic nature of genetic test results and underlying genetic heterogeneity, Holtzman et al. *(188)* predicted that these factors will become increasingly responsible for misinterpretations as genetic testing for "susceptibility alleles of low penetrance for common chronic diseases becomes more common." Reyna et al. *(169)* discussed the roles of incomplete penetrance and incorrect assumptions of the expected prevalence in the general population as sources of interpretation errors. These were both underlying causes of misinterpretation by ordering physicians when they attempted to interpret results of testing for mutations in the *APC* gene, which are responsible for familial adenomatous polyposis (FAP), a relatively rare, autosomal dominant form of colorectal cancer *(67)*. Nearly 20% of the 177 patients tested were inappropriate candidates, only 19% received

genetic counseling before testing, and only 17% of the patients provided written informed consent (an indirect measure of the degree to which testing was an informed decision). The lack of expertise in genetic principles led one-third of the ordering physicians to misunderstand that the failure to detect a mutation in an asymptomatic person could represent a false-negative result when a mutation had not been previously identified in the family. Because the penetrance of APC mutations is essentially 100%, this mistaken interpretation could lead to the inappropriate lack of surveillance for asymptomatic individuals. This example has widespread implications for molecular genetic testing, in general, as the vast majority of clinically available tests involve genes with moderate to extensive allelic heterogeneity and low *a priori* risks in the general population. The authors highlighted the need for genetic counseling prior to and following predictive testing to address these kinds of misinterpretations *(169)*.

6.3. IMPROVING THE ACCURACY OF TEST INTERPRETATION To bridge the gap between genetic test results and the ordering physician, several strategies have been proposed. Genetic counselors are increasingly becoming employed in testing laboratories to provide genetic expertise *(125,189)*. Increased regulatory guidelines and clinical and technical recommendations for molecular testing of several single-gene disorders have been published and are summarized by Amos and Patnaik *(19)*. Recommendations for the selection of appropriate candidates for genetic testing, the selection of appropriate genetic tests, provision of genetic counseling, and medical management for many disorders have been published *(149,182–185)* or can be found on the Internet (www.ashg.org/; www.nsgc.org/; www.geneclinics.org). Several authors have also cited the need for more comprehensive laboratory reports *(190)*.

6.4. PHYSICIAN TRAINING AND EXPERIENCE IN GENETIC DISORDERS IS LIMITED Because many genetic disorders are individually rare, training in their recognition and medical management as well as instruction in basic genetic principles are limited in medical schools (http://genetics.faseb.org/genetics/aphmg/aphmg1.htm). This lack of knowledge has implications for patient care, as providers who have less experience with a disorder are more pessimistic about prognosis, less aware of current treatment options, and are more likely to deliver information in a directive and nontherapeutic manner *(18)*. Increasingly, parents use the Internet to find comprehensive information about diagnostic criteria, potential treatments, opportunities to participate in clinical trials, and sources of support. In some cases, this results in the parents having more acquired expertise about the condition affecting the family than their local physician. In families with affected children, Internet use is correlated with the educational level of the parent and the presence of a specific diagnosis *(191–193)*.

6.5. UNEQUAL ACCESS TO GENETIC SERVICES As with other forms of health care, access to genetic services (especially in the United States) is unequal. Several factors are associated with this disparity, including educational and income levels *(103)*, social and cultural factors *(149,194)*, awareness by families and their physicians *(103,107,195)*, physicians' referral preferences *(107)*, insurance coverage *(196)*, and the limited number of genetic professionals. Although patients might be unsure of how genetic counseling could benefit them, most agree that they would prefer to be given the option. *(197)*. According to Hayflick et al. *(198)*, the main reason primary care physicians choose not to refer patients for genetic counseling is their perception that it would not benefit the family. These assumptions could be increasingly challenged as families become more knowledgeable about the benefits of genetic services through the media, especially the Internet.

7. ILLUSTRATIVE CASES IN GENETIC COUNSELING

7.1. CASE 1: DIAGNOSTIC TESTING FOR CYSTIC FIBROSIS A 9-mo-old child has a history of chronic respiratory infections, and foul-smelling stools and is small for her age. Testing of the CFTR gene shows that she is homozygous for the ΔF508 mutation causing cystic fibrosis (CF). Her older sister, age 4, also has a history of respiratory infections and measures less than the 5th percentile for height and weight. Blood is obtained from her and tested. The results show that she is also homozygous for the ΔF508 mutation.

The decision whether to test an older sibling for the ΔF508 mutation is not controversial. This is especially true if the child is symptomatic, there is a positive family history, and the testing is diagnostic, as in the example case provided. Although not without challenges in determining the most effective and compassionate way to communicate information about the diagnosis, prognosis, and inheritance of a disorder and anticipating and addressing the psychological implications for the child and the family *(199)*, genetic testing that guides pertinent medical management is considered standard of care. The autosomal inheritance of CF, the 25% recurrence risk for the parents, the availability of prenatal diagnosis, and other reproductive options (such as artificial insemination by donor and preimplantation diagnosis) are pertinent issues but are not likely to be burning ones for the parents at this time. Rather, it is the psychological impact of having both children diagnosed within a short time of each other that is paramount.

7.1.1. Delayed Diagnosis: The Option of Newborn Screening Because of the variability in phenotypes, some children might not be diagnosed until past the expected age of disease onset. In response, there has been some support to include disorders such as CF in newborn screening programs *(145)*, although some argue that the case for early detection and treatment has not been made conclusively *(200)*. Other genetic disorders are also being studied to determine whether they would be appropriate candidates for newborn screening *(201,202)*. Arguments for inclusion include variable benefit of early treatment, parental knowledge of the recurrence risks for future pregnancies, and avoidance of unnecessary tests and evaluations. Arguments against inclusion include inadequate knowledge of the natural history of the disorder, the economic costs (especially in cases in which the false-positive rate is high), and the lack of a direct relationship between treatment and disease. Neonatal testing for PKU is considered the gold standard by which the others are measured. As Burke noted *(203)*, even in the case of PKU, the true costs (including both economic and psychosocial ones) have not been carefully assessed. Nonetheless, the benefits of this testing are clear. In contrast, once screening for disorders with less effective treatments are

begun, a careful accounting of all of the costs and benefits will be necessary to determine if testing is warranted.

7.2. CASE 2: FRAGILE X DIAGNOSTIC AND CARRIER TESTING A 9-yr- old girl has developmental delay and autistic behaviors. In addition to other genetic testing, analysis of the FMR1 gene is ordered and shows expansion of the CGG repeat of over 1000. Her mother wonders if the mental retardation in her former husband's family is related to this diagnosis of Fragile X syndrome. The mother is also concerned about her former husband's daughter, an 18-yr-old girl, who has some mild cognitive delays and is sexually active.

This case illustrates how diagnostic genetic testing has implications for the relatives of an affected individual. The younger daughter was an appropriate candidate for Fragile X testing because the result explained the presence of her abnormal phenotype. This diagnosis can result in more appropriate educational placements and possibly enhanced eligibility for services. The implications of the results of her testing for others in the family create the complexity. The 18-yr-old paternal half-sister has some mild cognitive delays and there could be others in the family with mental retardation. In the absence of the documented mutation in a relative, it is unlikely that the older girl would be offered Fragile X testing because of her mild and nonspecific phenotype. Ironically, if she were to become pregnant, this offer would more likely be forthcoming, as it would increase her chance of being seen by a health care professional with more knowledge in genetics *(204)*. Screening women (and sometimes their partners) during pregnancy for carrier status of several genetic conditions (even in the absence of family history) is currently recommended. As Collins noted, the detection of carrier status during pregnancy is less than ideal, as it removes many reproductive options from consideration *(109)*.

7.2.1. Knowledge of the Family History Can Determine Appropriate Testing Candidates The diagnosis of Fragile X syndrome in the younger sister could prompt carrier testing of the 18-yr-old paternal half-sister, but her result is unlikely to show an expanded repeat. The careful delineation of the family history is an important factor in this conclusion. Because the term "mental retardation" is a general one, careful questioning could reveal that the father's family history consists of an individual with Down's syndrome. The presence of a normal phenotype in the girls' father does not exclude the possibility that he has a premutation but remains unaffected. In this case, all of his daughters would inherit this mutation, but it is very unlikely to expand into a range that would be associated with learning problems in a female. All of his sons would inherit his Y chromosome and not be at risk *(205)*.

7.2.2. Carrier Testing in Adolescents The decision whether to offer testing to the older paternal half-sister is a complicated one. Testing the younger daughter's mother is likely to reveal that she is the parent from whom the Fragile X mutation is derived. Regardless of this result, testing the 18-yr-old daughter might be somewhat informative, although the presence of a normal FMR1 pattern does not determine the cause of her learning problems and does not eliminate the chance that it could have a genetic etiology with potential risks to future children. Because of her age, this testing should only proceed after she has given informed consent so that she understands the potential implications *(45,206)*. This process might prevent testing from occurring, as she may not want to be tested for various reasons, including psychological grounds *(31,41–43,95,98)* or financial considerations *(207)*. Alternatively, she might live somewhere distant and communications with her about this topic might be mediated by the family's assumptions and limited understandings of the risk *(42,59,208,209)*. Direct disclosure of genetic information to relatives without the consent of the individual (or in this case, the younger daughter's mother) by genetic professionals is not generally accepted practice *(57,61)*. Efforts to communicate these issues to her could include sending a letter that explains her potential risk without revealing specific information about the genetic status of other relatives and providing contact information for a genetic professional in her area. As the chance for inheriting a genetic change related to a clear disease risk, against which specific actions could be taken to prevent disease development, increases, the role of genetic professionals in the disclosure become more acceptable *(65)*. However, in practice, it remains rare for genetic professionals to breach this confidentiality *(60,62–64)*.

7.2.3. Implications for Extended Family Members The diagnosis of Fragile X syndrome in a proband has profound implications for many other relatives because of the nature of the mutation *(210)*. It would be expected that additional relatives would be at risk for being carriers of either a premutation or, in some instances, a full mutation. The variable expression of the phenotype between males and females and between those with different sized expansions, that can be present in differing proportions of relevant tissues, increases the complexity of the genetic information. Careful delineation of the family history can determine appropriate candidates for testing. The subject of asymptomatic carrier testing in children under 18 yr of age requires special considerations *(31,36,45,46,49,199,210)*. The complexity of the genetic information and potential implications for many relatives makes disclosure efforts challenging.

7.3. CASE 3: DIAGNOSTIC TESTING FOR GJB2 (CONNEXIN 26) A 3-yr-old boy with nonsyndromic, profound, sensorineural hearing loss has genetic testing that shows he is homozygous for the 35delG mutation in the GJB2 gene. His parents want to know how their son can have a genetic condition when no one else in the family is deaf and what the chance is that his children will be deaf. Their 8-mo-old daughter passed her newborn hearing screen, but the parents are still concerned and request genetic testing to determine her mutation status. They also want to know whether, if they were to have another child, prenatal diagnosis would be available.

One variation of the diagnostic use of genetic testing is illustrated by cases in which an asymptomatic person is tested because he or she belongs to a family in which a relative has a genetic condition for which the mutation is known. The acceptance of this testing can vary depending on the occurrence of reliable clinical features, enabling the condition to be detected and treated before irreversible harm occurs. The detection of hearing loss in infants can be variable depending on diagnostic techniques employed and the degree of hearing loss. Targeting diagnostic testing of relatives known to be at elevated risk for a genetic disease is not new and increases the reliable detection

of affected individuals. This is the rationale behind prenatal genetic testing of a fetus because of a family history.

Because of its ability to detect carriers, as well as affected individuals, genetic testing has the potential to provide not only the information that is sought (affected vs unaffected) but simultaneously provides additional information that is irrelevant to the current and future health of the individual (carrier status) and becomes relevant only (if at all) at an age when reproductive decisions are made. The comprehensive answers that genetic testing provides has the potential to create conflicts. The answer to a specific clinical question comes inextricably encased in information that answers questions that were not asked (and are not clinically helpful at that time) about individuals who have no clinical symptoms. In the absence of a proven benefit, genetic testing of an asymptomatic child, because of the irretrievable loss of their autonomy, is usually discouraged. When the testing result of the proband is difficult to interpret, testing collateral relatives becomes even more problematic. If the proband had been a compound heterozygote with one well-known mutation (such as 35delG) but with a second allele that had a change of unknown significance, testing his sister would be more likely to result in ambiguity, especially if she inherited the same genotype. The expansion of genetic testing options to detect conditions with variable natural histories is likely to result in testing an increasing number of asymptomatic children.

7.3.1. Expanded Use of Genetic Testing for Hearing Loss Genetic testing for congenital hearing loss has expanded as the genetic etiologies of many types of sensorineural hearing loss have been determined in the past 10 yr *(81,211–215)*. Reviews of the clinical features, diagnostic work-ups, and molecular diagnoses of the major forms of syndromic and nonsyndromic hearing loss are available *(81,216)*. Syndromic hearing loss is associated with the occurrence and/or risk of occurrence of other features, whereas nonsyndromic implies the lack of such features. Additional subclassifications include the timing, the degree, and the type of loss. The usual prevalence stated for congenital hearing loss of greater than 35 dB ranges between 1 to 2 in 1000 *(215)*, increasing to 3.7–56.3 per 1000 if later onset hearing loss is considered. Approximately 50–60% of all cases are considered to have genetic etiologies. Of the genetic forms, 75–80% are inherited in an autosomal recessive pattern, characterized by profound, prelingual deafness *(81)*, 50% of which are caused by mutations in the GJB2 genes *(217)*. As many as 70% of GJB2-related cases are the result of a single mutation, 35delG *(218)*, although the 167delT mutation is the most common in Ashkenazi Jews. The carrier frequency of 35delG is estimated to be between 2% and 3% in Caucasians *(219)*. Carrier frequencies for the 35delG mutation among other ethnicities have not been well established.

7.3.2. Newborn Screening for Hearing Loss The evolution of newborn hearing screening has been described *(81)* and detects 2–3 babies per 1000 with permanent hearing loss. Prior to screening, the average age of diagnosis for severe hearing loss was 30 mo of age, with greater delay for mild and moderate losses *(184)*. Newborn screening has resulted in early detection and intervention that allows infants to develop verbal language at a comparable rate to those without loss *(220,221)*. Children with GJB2-related deafness are particularly good candidates for cochlear implants because of the lack of associated neurological involvement *(217)*, although this intervention is somewhat controversial in the Deaf community. Because of the autosomal recessive inheritance of GJB2-related deafness, the overwhelming majority, of affected children, approx 90% are born to hearing parents *(81)*.

7.3.3. Interest in Genetic Testing for Hearing Loss Varies Reported interest in genetic testing and actual uptake can vary considerably. Parents of children with hearing loss reported high levels of interest in genetic testing but poor knowledge about the meaning of their child's result *(222)*. The potential use of prenatal testing generated concern when an article by Middleton et al. *(223)* indicated that some individuals who consider themselves as culturally Deaf indicated they would use prenatal diagnosis and prefered to have a deaf child. Careful examination of the data show that this response was chosen by only four participants, although it is the proportion (29%) that is most widely cited *(224,225)*. However, the majority of the total sample population, ascertained at a conference about Deaf culture, did not express a preference. Others have also reported projected decisions about terminating a pregnancy in which deafness was prenatally detected *(225)*. A survey of a wider sample of individual with hearing loss showed large variability in acceptance of testing and prenatal diagnosis *(226)*. The experience in Greece when prenatal testing was offered to couples with no family history of deafness illustrates how the rate of testing uptake is influenced by the manner in which testing is offered and by cultural factors *(224)*. The underlying motivations for genetic testing of children with hearing loss in Israeli Jewish parents varied with their religious preference *(227)*. A culturally sensitive approach to genetic counseling, regardless of the culture or the disorder involved, is warranted *(101)* and includes the use of value-neutral language (the use of the word *chance* rather than *risk*, for example) and the use of appropriately trained interpreters.

7.4. CASE 4: PREDICTIVE AND PRENATAL TESTING FOR HUNTINGTON'S DISEASE A 32-yr-old man is told by his older sister that his biological father, who left the home when the man was a boy, recently died of Huntington's disease (HD) that was confirmed by DNA testing. His sister also tells him that they had different fathers so she is not at risk for HD. The man's wife is currently 14 wk pregnant with their first child. The couple has decided that they would like to find out the genetic status of the baby, but they do not want to know whether the man inherited the altered HD allele.

A large portion of the research resulting in the development of genetic counseling protocols for presymptomatic genetic testing and disclosure of test results has been informed by the experiences of predictive and susceptibility testing, especially for HD, hereditary breast and ovarian cancer, and other hereditary cancers. It is in these cases, rather than in the diagnostic arena, that issues such as the voluntary nature of the test, the necessity for informed consent, and concerns about potential misuse of genetic information are proposed and debated *(13,14,177,228–234)*.

7.4.1. Interest in Predictive and Susceptibilty Testing Surveys of at-risk or other populations have been reported that

attempt to measure awareness and interest in various forms of predictive testing *(73,92, 235–247)*, although, invariably, the proportion of individuals who have expressed actually proceed with testing is much lower than the proportion of individuals who have expressed an interest. In HD, for example, the uptake is estimated to be between 10% and 15% *(248)*, but as many as 60–75% of some at-risk populations have expressed interest *(228, 233)*. A correlation between rates of uptake and the availability of treatment has been noted, with higher uptakes for hereditary breast and ovarian cancer (approx 50%) and for FAP (approx 80%) *(236)*. Although expressed interest in testing cannot be used as a valid predictor of testing uptake, it has been a common topic in the literature, and various socioeconomic, demographic, and psychological correlations have been noted.

7.4.2. Factors That Affect the Uptake of Testing Several factors increase participation in genetic testing. These factors include (1) testing offered concurrently with an information session *(249)*, (2) testing at home instead of in a clinical setting *(250)*, and (3) testing based on a sample that can be obtained easily *(251)*. The manner in which it is offered also affects use *(243,252–254)*. Marteau and Croyle *(236)* noted the importance of determining whether high uptake of testing is desired before attempting to increase participation. Protocols for predictive testing typically allow varying amounts of time to elapse between the initial appointment and the procurement of the sample *(255)*, resulting in a skewed population that is finally tested. This self-selection process *(256)* is probably responsible, along with comprehensive assessment and counseling, for the relatively low rates of major long-term psychological problems measured after predictive testing. Their presence seems to be more strongly related to pretesting psychological functioning than to whether a mutation is detected *(49,248,257,258)*. Several demographic characteristics might play a role. Knowledge of the disorder because of a family history usually acts to moderate reactions following testing because previous experience with the disorder is correlated with rating it as less serious *(253,257)*. This phenomenon has implications for genetic counseling accompanying population screening, including expanding the disorders included in newborn screening *(145)*. Psychological factors such as increased perception of risk *(93,259)*, levels of worry *(260)*, and the need to reduce uncertainty *(261)* have also been reported to increase test usage.

In general, candidates for predictive and susceptibility testing overestimate their risks as well as the benefits of testing while underestimating the potential risks and limitations associated with testing. Increased information about these factors, as occurs during a genetic counseling session, is associated with decreased interest *(127)*, a decreased perception of risk *(258)*, especially in women with lower education levels *(180)*, and an increased hesitancy to participate in research *(251)*.

7.4.3. Direct Testing for HD An exhaustive list of publications have been devoted to HD testing *(262–268)*. It can be accomplished by direct testing of the individual (the consultand) requesting the test. The causative mutations, various degrees of expansion of a CAG repeat in the gene, are detectable by current molecular techniques *(269)*. If the consultand in the presented case had requested predictive testing, it would be prudent to obtain a copy of his father's DNA testing results to assure their accuracy. The detection of an expansion beyond the normal threshold in the consultand is very sensitive and specific for the eventual occurrence of symptoms of HD, although there is some decreased penetrance when the expansion is close to the normal range *(270)*. Fathers tend to contribute larger expansions *(271)*. The size of the expansion has general predictive qualities, with larger repeats associated with an earlier onset, but specific predictions for individuals based on their repeat size cannot be directly made *(272)*. Although much of the literature is devoted to presymptomatic testing, symptomatic individuals are candidates for testing although issues of informed consent and patient autonomy are complicated by the neurological deterioration that characterizes the disorder and the greatly increased risk for suicide in the early stages *(273)*.

7.4.4. Linkage Testing The availability of direct testing has removed some of the logistical barriers that were present during the intervening period from the identification of the first set of linked markers until the isolation of the HD gene *(257,263,274–276)*. Linkage may be used in testing for disorders for which there are no common mutations, the proband does not have an identifiable mutation, or the gene has not been cloned. The requirement that blood be obtained from several key relatives in order to establish the linkage phase can disqualify many testing candidates because of strained family relationships, uncooperative relatives, and deaths of key affected individuals *(277)*. However, linkage testing has the distinct advantage of uncoupling the results of a parent with those of a child and has been used in prenatal diagnosis.

7.4.5. Testing Decisions in HD Huntington's disease, by virtue of its devastating phenotype and the early identification of linked markers (in 1986) allowing it to be one of the first examples of accurate predictive testing, has served as a prototype for the ethical and social issues that arise for conditions with similar attributes. Extensive protocols have been developed for HD *(278–281)*. Bird *(273)* described the essential features of the phenotype that make the decision to have predictive testing so difficult: "… cognitive impairment in prime of life, socially embarrassing movements, serious cognitive and behavioral changes, shortened life span, no cure or treatment, and a 50% risk for each child to inherit the gene." When individuals receive a positive predictive result, it serves to "… prolong an intolerable ambiguity …" *(273)* because the uncertainty surrounding whether or not they have inherited the expansion is replaced with the ambiguity regarding the timing of the onset of symptoms. Ambiguous results (as in the early linkage testing) were associated, somewhat unexpectedly, with more distress than either an increased or decreased risk. Butow et al. *(258)* described a similar phenomenon in participants who declined to receive their BRCA testing results. The complexities of these responses, including the adverse psychological reactions of some who were mutation negative *(282)*, underscores the need for predictive testing to be offered in conjunction with comprehensive counseling by experienced centers.

7.4.6. Long-Term Psychological Impact of Predictive Testing for HD Overall, participants in these highly regulated and standardized counseling and testing protocols faired

better than expected *(93,248)*, although subsets of patients did have problems. The presence of psychiatric or emotional problems prior to testing appeared to be more predictive of which participants would experience significant and long-term distress, rather than the results of the testing. This was observed in both HD and familial cancer *(93,257)*. Adverse reactions, as defined by trouble coping with the result, were more common for those who had made irreversible decisions based on their assumptions about their genetic status. One of the most concerning effects, especially for HD, was the potential for testing to result in an increased suicide rate. Almqvist *(283)* studied catastrophic events (defined as attempted or completed suicide or hospitalization) among 4524 participants at 100 HD testing centers in 21 countries. This study found the rate of catastrophic events among tested patients was similar to that of individuals affected by HD (5.7%) *(273,280)*, or 7–10 times the average rate in the United States. All of the completed suicides were in mutation-positive individuals who had begun to show clinical signs. Seven individuals (0.3%) receiving a decreased risk also experienced a catastrophic event. Although the profile of someone at highest risk, seems to be an unemployed woman with a previous psychiatric history who is mutation positive and develops symptoms less than 1 yr after testing, all of those who completed a suicide were men.

7.4.7. Exclusion and Nondisclosing Prenatal Testing The decision to have predictive testing is complex and difficult and one that is not optimally undertaken when the individual or his or her partner is pregnant. Some parents who have a 50% risk because they have an affected parent might not want to have their genetic status revealed but still request that the fetal status be determined. Direct fetal testing of the trinucleotide repeat (when an expansion has been inherited) would simultaneously reveal the at-risk parent's status. Exclusion testing determines which grandpaternal allele the fetus has inherited, preserving the parental at-risk status. This testing determines whether the fetus is at 50% risk (like the parent) or not at risk. In the case example, the at-risk father has a 50% chance that the allele he inherited from his affected father has an expansion. The at-risk man has also inherited a normal allele from his unaffected mother. Likewise, the fetus will inherit one allele from the (at-risk) father and one from the unaffected mother. If the fetus inherits the allele that the at-risk man received from his affected father, there is a 50% risk. Alternatively, if the testing shows fetal inheritance of the paternal *grandmother's* allele, the risk is essentially zero *(284)*. Although first described in 1985 *(285)*, exclusion testing can now be performed in conjunction with preimplantation genetic diagnosis (PGD). A variation, called nondisclosing prenatal testing, has been done in which embryos are analyzed for the repeat expansion and only those with the normal pattern are transferred *(286)*. This procedure does not reveal the at-risk person's status to them, but if an expansion is detected, it will be revealed to some of the laboratory personnel. The practical and ethical issues, including the necessity for undergoing repeated cycles of in vitro fertilization and PGD even when the at-risk parent did not inherit the expansion are discussed by Sermon et al. *(284)*.

7.5. CASE 5. GENETIC TESTING FOR INHERITED CANCER SUSCEPTIBILITY A 45-yr-old woman's mother died of breast cancer when she was in her fifties. The woman's maternal aunt had breast cancer diagnosed when she was in her late forties and then later died of ovarian cancer. The woman is very concerned about her risk of developing breast cancer and wants to have genetic testing performed. She reads about a commercial laboratory offering the test and convinces her local doctor to order it. She pays for it herself, as she does not want her insurance company to be notified. Her test result shows a change in BRCA1 of "unknown clinical significance." Although her doctor is reassuring, the woman tells her sister that a mutation has been found and advises her to be tested too.

Burke observed that genetic tests are increasingly being offered before a clear understanding of their benefits and costs, including the potential efficacy of interventions, have been accurately assessed *(203)*. The increasing development of genetic tests with low predictive values, as in those for susceptibility for diseases, require that this cost-benefit determination be made, where possible, to "… help clinicians integrate genetic testing into general clinical practice …" with the help of standardized criteria. Arguing that this accounting should "… precede the introduction of a genetic test …," the true measure of its success would be the extent to which it meets a "… well-defined clinical problem for which (it is) expected to improve care …" *(203)*.

Much has been written about the future of genetic testing with the potential to detect susceptibility alleles that could act as targets for prevention or treatment. The most recent literature is beginning to reflect these expectations and measures the extent to which genetic testing influences (or does not influence) subsequent behaviors *(287–289)*. Much remains to be learned. The direct marketing of genetic tests concerns most genetic professionals because of the potential for misinterpretation of both positive and negative results *(290)*. These misinterpretations have the potential to be magnified in families because of the implications for relatives. Prior assumptions about the level to which genes are deterministic of health *(17,76)* could cause those with a positive result to become fatalistic. Research is needed to determine the most effective and practical genetic counseling strategies for an increasingly large number of people who will be offered genetic testing for treatable conditions *(145)*.

The public's use of other alternative sources of genetic information, like that found on the Internet, is expected to continue to increase, although some of the available information is misleading or inaccurate. For example, when pamphlets describing Down's syndrome *(291)* and cystic fibrosis *(148)* were analyzed, it was concluded that the descriptions of Down's syndrome were more negatively biased. Those prepared by a commercial laboratory described a more pessimistic view than those written by one in an academic setting.

There is an increasing interest in the use of family histories in general medical practice to stratify families by risk *(111)*. The range of reliability and completeness of the information will be important to determine. This tool has been historically underutilized by family physicians. The presence of a strong family history that could indicate a genetic predisposition to a chronic disease may not be detected by the family's physician. Acheson et al. reported that family practitioners discuss the family history in only about half of the new-patient visits

and only one-quarter of the time in established-patient visits. This discussion of the family history lasted, on average, 2.5 min *(131)*. The use of a touchscreen computer to collect a family history of cancer was evaluated *(152)*, with over 1400 participants over approx 1 yr. The information collected was interpreted by a genetics professional who provided a basic cancer risk assessment in broad terms and surveillance recommendations within 10 d by a tailored letter. The ability of computers to extend the range of practice of genetic professionals is in its infancy.

Finally, it is interesting to read the article by Michie and colleagues *(290)* on why negative results sometimes fail to reassure patients, for a broader understanding of some of the misconceptions about genes and genetic tests that are held by the public. Concern has been expressed that a negative result might falsely reassure, as was shown in FAP example *(67)*. Although Johnson et al. *(292)* reported that those testing positive for the APC mutation were more likely to adhere to recommended screening protocols, those who tested negative for the familial mutation were not falsely reassured. Michie et al. *(290)* found that one-third of those testing negative for a familial APC mutation desired continued surveillance. Individuals who had received their results from a nongenetic health professional were more likely to request additional screening. Interviews with nine members of the family indicated several misunderstandings about genetic testing; most notably, some who did not inherit the mutation still perceived themselves at risk and thought that genetic status was fluid and able to change in the future, leading to the conclusion that the test could not be trusted. The abstract nature of using blood to detect cancer of the bowel was also not appreciated. Other family myths also became part of the interpretation.

REFERENCES

1. Donnai, D. Genetic services. *Clin. Genet*. **61:**1–6, 2002.
2. Wilcke, J. T. R., Seersholm N., Kok-Jensen, A., and Dirksen, A. Transmitting genetic risk information in families: attitudes about disclosing the identity of relatives. *Am. J. Hum. Genet*. **65:**902–909, 1999.
3. Wright, C., Kerzin-Storrar, L., Williamson, P., et al. Comparison of genetic services with and without genetic registers: knowledge, adjustment, and attitudes about genetic counselling among probands referred to three genetic clinics. *J. Med. Genet*. **39:**e84, 2003.
4. Jorde, L., Carey, J., Bamshad, M., and White, R., eds. *Medical Genetics,* 3rd ed. Mosby, St Louis, MO, 2003.
5. Nussbaum, R., McInnes, R., and Willard, H., eds. *Thompson and Thompson Genetics in Medicine,* 6th ed., W. B. Saunders, Philadelphia, PA, 2001.
6. Botkin, J. R., McMahon, W. M., Smith, K. R., and Nash, J. E. Privacy and confidentiality in the publication of pedigrees—a survey of investigators and biomedical journals. *J. Am. Med. Assoc*. **279:**1808–1812, 1998.
7. Byers, P. and Ashkenas, J. Pedigrees—publish? or perish the thought. *Am. J. Hum. Genet.* **63:**678–681, 1998.
8. Dickens, B., Pei, N., and Taylor, K. Legal and ethical issues in genetic testing and counseling for susceptibility to breast, ovarian and colon cancer. *Can. Med. Assoc. J*. **154:**813–818, 1996.
9. Esplen, M. J., Madlensky, L., Butler, K., et al. Motivations and psychosocial impact of genetic testing for HNPCC. *Am. J. Med*. **103:**9–15, 2001.
10. Evans, D., Maher, E., Maclead, R., Davies, D., and Craufurd, D. Uptake of genetic testing for cancer predisposition—ethical issues. *J. Med. Genet*. **34:**746–749, 1997.
11. Hubbard, R. and Lewontin, R. C. Pitfalls of genetic testing. *N. Engl. J. Med*. **334:**1192–1194, 1996.
12. Lerman, C., Hughes, C., Lemon, S. J., et al. What you don't know can hurt you: adverse psychologic effects in members of BRCA1-linked and BRCA2-linked families who decline genetic testing. *J. Clin. Oncol*. **16:**1650–1654, 1998.
13. Billings, P. R., John, M. A., Decuevas, M., Beckwith, J., Alper, J. S., and Natowicz, M. R. Discrimination as a consequence of genetic testing. *Am. J. Hum. Genet*. **50:**476–482, 1992.
14. Hahn, M., Saeger, H., and Sackert, H. Hereditary colon cancer: clinic consequences of predictive molecular testing. *Int. J. Colorectal Dis*. **14:**184–193, 1999.
15. Lubinsky, M. Bearing bad news: dealing with the mimics of denial. *J. Genet. Counsel.* **3:**5–11, 1994.
16. Cary, J. Editor's note. *Am. J. Med. Genet*. **110:**1–2, 2002.
17. Chapple, A., May, C., and Campion, P. Lay understanding of genetic disease: a British study of families attending a genetic counseling service. *J. Genet. Counsel*. **4:**281–301, 1995.
18. Hill, V., Sahhar, M., Aitken, M., Savarirayan, R., and Metcalfe, S. Experiences at the time of diagnosis of parents who have a child with a bone dysplasia resulting in short stature. *Am. J. Med. Genet.* **122A:**100–107, 2003.
19. Amos J. and Patnaik M. Commercial molecular diagnostics in the US: The human genome project to the clinical laboratory. *Hum. Mutat.* **19:**324–333, 2002.
20. Burke, W., Pinsky, L., and Press, N. Categorizing genetic tests to identify their ethical, legal, and social implications. *Am. J. Med. Genet.* **106:**233–240, 2001.
21. Pagon, R. Genetic testing for disease susceptibilities: consequences for genetic counseling. *Trends Mol. Med*. **8:**306–307, 2002.
22. Green, J. Health care evaluation and cost analysis of clinical and genetic screening for von Hippel Lindau in Newfoundland. *Am. J. Hum. Genet*. **57:**A95, 1995.
23. Stanley, A., Gaff, C., Aittomaki, A., Fabre, L., Macrae, F., and St. John, D. J. Value of predictive genetic testing in management of hereditary non-polyposis colorectal cancer (HNPCC). *Med. J. Aust.* **172:**313–316, 2000.
24. Kodish, E., Wiesner, G., Mehlman, M., and Murray, T. Genetic testing for cancer risk, how to reconcile the conflicts. *J. Am. Med. Assoc*. **279:**179–181, 1998.
25. Bell, J. The new genetics in clinical practice. *Br. Med. J*. **316:**618–620, 1998.
26. Collins, F. and Guttmacher, A. Genetics moves into the mainstream. *J. Am. Med. Assoc*. **286:**2322–2324, 2001.
27. Khoury, M., Burke, W., and Thomson, E. Genetics and public health: a framework for the integration of human genetics into public health, in *Genetics and Public Health in the 21st Century*, Khoury, M. J., Burke, W., and Thomson, E., eds. Oxford University Press, New York, 2001.
28. Wulfsberg, E. The impact of genetic testing on primary care: where's the beef. *Am. Fam. Phys*. **61:**971–978, 2000.
29. Smith, K., West, J. A., Croyle, R., and Botkin, J. Familial context of genetic testing for cancer susceptibility: moderating effect of siblings' test results on psychological distress one to two weeks after BRCA1 mutation testing. *Cancer Epidemiol Biomarkers Prev*. **8:**385–392, 1999.
30. Geller, G., Doksum, T., Bernhardt, B. A., and Metz, S. Participation in breast cancer susceptibility testing protocols: Influence of recruitment source, altruism, and family involvement on women's decisions. *Cancer Epidemiol Biomarkers Prev*. **8:**377–383, 1999.
31. Wertz, D. C., Fanos, J. H., and Reilly, P. R. Genetic testing for children and adolescents who decides? *J. Am. Med. Assoc*. **272:**875–881, 1994.
32. Grosfeld, F. J. M., Lips, C. J. M., Beemer, F. A., van Spijker, H. G., Brouwers-Smalbraak, G. J., and ten Kroode, H. F. J. Psychological risks of genetically testing children for a hereditary cancer syndrome. *Patient Educ. Counsel*. **32:**63–67, 1997.
33. Grosfeld, F., Beemer, F., Lips, C., Hendricks, K., and ten Kroode, H. Parents' responses to disclosure of genetic test results of their children. *Am. J. Med. Genet.* **94:**316–323, 2000.
34. Codori, A., Petersen, G. M., Boyd, P. A., Brandt, J., and Giardiello, F. M. Genetic testing for cancer in children: short term psychological effects. *Arch. Pediatr. Adolesc. Med*. **150:**1131–1138, 1996.

35. Codori, A. M., Zawacki, K., Petersen, G., et al. Genetic testing for hereditary colorectal cancer in children: long-term psychological effects. *Am. J. Med. Genet*. **116A:**117–128, 2003.
36. Fryer, A. Inappropriate genetic testing of children. *Arch. Dis. Child*. **83:**83–85, 2000.
37. Jarvinen, O., Aalto, A.-M., Lehesjoki, A-E., Lindlof, M., Soderling, A., and Kaariaines, H. Carrier testing of children for two X-linked diseases in a family based setting: a retrospective long term psychosocial evaluation. *J. Med. Genet*. **36:**615–620, 1999.
38. Jarvinen, O., Lehesjoki, A., Lindlof, M., Uutela, A., and Kaariainen, H. Carrier testing of children for two X-linked diseases: a retrospective study of the test results and social and psychological significance of the testing. *Pediatrics* **106:**1460–1465, 2000.
39. Fanos, J. and Johnson, J. Barriers to carrier testing for adult cystic fibrosis sibs: the importance of not knowing. *Am. J. Med. Genet*. **59:**85–91, 1995.
40. Fanos, J. and Gatti, R. A mark on the arm: misconceptions of sibs and carrier status and genetics of AT. *Am. J. Med. Genet*. **86:**338–346, 1999.
41. McConkie-Rosell, A., Spiridigliozzi, G., Iafolla, T., Tarleton, J., and Lachiewicz, A. Carrier testing in the fragile X syndrome: attitudes and opinions of obligate carriers. *Am. J. Med. Genet*. **68:**62–69, 1997.
42. McConkie-Rosell, A., Spiridigliozzi, G., Sullivan, J., Dawson, D., and Lachiewicz, A. Carrier testing in fragile X syndrome: effect on self-concept. *Am. J. Med. Genet*. **92:**336–342, 2000.
43. James, C., Holtzman, N., and Hadley, D. Perceptions of reproductive risk and carrier testing among adolescent sisters of males with chronic granulomatous disease. *Am. J. Med. Genet*. **119C:**60–69, 2003.
44. American Academy of Pediatrics, Committee on Bioethics Ethical issues with genetic testing in pediatrics. *Pediatrics* **107:**1451–1455, 2001.
45. ASHG/ACMG Report. Points to consider: ethical, legal, and psychosocial implications of genetic testing in children and adolescents. *Am. J. Hum. Genet*. **57:**1233–1241, 1995.
46. Ross, L. and Moon, M. Ethical Issues in genetic testing of children. *Arch. Pediatr. Adolesc. Med.* **154:**873–879, 2000.
47. Elger, B. and Harding, T. Testing adolescents for a hereditary breast cancer gene. Respecting their autonomy is in their best interests. *Arch. Pediatr. Adolesc. Med*. **154:**113–119, 2000.
48. Varekamp, I., Suurmeijer, T. P., Brocker-Vriends, A. H., et al. Carrier testing and prenatal diagnosis for hemophilia: experiences and attitudes of 549 potential and obligate carriers. *Am. J. Med. Genet.* **37:**147–154, 1990.
49. Michie, S., Bobrow, M., Marteau, T. M. and the FAP Collaborative Research Group. Predictive genetic testing in children and adults: a study of emotional impact. *J. Med. Genet*. **38:**519–526, 2001.
50. Ayme, S, Ayme, S., Macquart-Moulin, G., Julian-Reynier, C., Chabal, F., and Giraud, F. Diffusion of information about genetic risk within families. *Neuromusc. Disord.* **3:**571–574, 1994.
51. McConkie-Rosell, A., Robinson, H., Wake, S., Staley, L., Heller, K., and Cronister, A. Educating extended family members about the inheritance of the fragile X syndrome. *Dev. Brain Dysfunct*. **8:**390–395, 1995.
52. Julian-Reynier C., Eisinger, F., and Vennin, P. Attitudes towards cancer predictive testing and transmission of information to the family. *J. Med. Genet*. **33:**731–736, 1996.
53. Julian-Reynier, C. Disclosure to the family of breast/ovarian cancer genetic test results: patient's willingness and associated factors. *Am. J. Med. Genet*. **94:**13–18, 2000.
54. Hughes, C., Lynch, H., Durham, C., et al. Communication of BRCA1/2 test results in hereditary breast cancer families. *Cancer Res. Ther. Control* **8:**51–59, 1999.
55. Claes, E., Evers-Kiebooms, G., Boogaerts, A., Decruyenaere, M., Denayer, L., and Legius, E. Communication with close and distant relatives in the context of genetic testing for hereditary breast and ovarian cancer in cancer patients. *Am. J. Med. Genet.* **116A:**11–19, 2003.
56. Peterson, S., Watts, B., Koehly, L., et al. How families communicate about HNPCC genetic testing: findings from a qualitative study. *Am. J. Med. Genet*. **119C:**78–86, 2003.
57. ASHG Social Issues Subcommitttee on Family Disclosure. ASHG statement: Professional disclosure of familial genetic information. *Am. J. Hum. Genet*. **62:**474–483, 1998.
58. Benkendorf, J., Reutenauer, J., Hughes, C., et al. Patients' attitudes about autonomy and confidentiality in genetic testing for breast–ovarian cancer susceptibility. *Am. J. Med. Genet.* **73:**296–303, 1997.
59. Bernard, L., McGillivray, B., Van Allen, M., Friedman, J., and Langlois, S. Duty to recontact: a study of families at risk for fragile X. *J. Genet. Counsel*. **8:**3–15, 1999.
60. Deftos, L. The evolving duty to disclose the presence of genetic disease to relatives. *Acad. Med*. **73:**962–968, 1998.
61. Dugan, R. B., Wiesner, G. L., Juengst, E. T., O'Riordan M. A., Matthews, A. L., and Robin, N. H. Duty to warn at-risk relatives for genetic disease: genetic counselors' clinical experience. *Am. J. Med. Genet*. **119C:**27–34, 2003.
62. Falk, M., Dugan, R., O'Riordan, M. A., Matthews, A., and Robin, N. Medical geneticists' duty to warn at-risk relatives for genetic disease. *Am. J. Med. Genet*. **120A:**374–380, 2003.
63. Fitzpatrick, J., Hahn, C., Costa, T., and Huggins, M. The duty to recontact: attitudes of genetic service providers. *Am. J. Hum. Genet*. **64:**852–860, 1999.
64. Lehmann, L., Weeks, J., Klar, N., Biener, L., and Garber, J. Disclosure of familial genetic information: Perceptions of the duty to inform. *Am. J. Med*. **109:**705–711, 2000.
65. Pelias, M. Duty to disclose: a legal perspective. *Am. J. Med. Genet*. **39:**347–354, 1991.
66. Hoffman, K. J., Tambor, E. S., Chase, G. A., Geller, G., Faden, R. R., and Holtzman, N. A. Physicians' knowledge of genetics and genetic tests. *Acad. Med*. **68:**625–632, 1993.
67. Giardiello, F. M., Brensinger, J. D., Petersen, G. M., et al. The use and interpretation of commercial APC gene testing for familial adenomatous polyposis. *N. Engl. J. Med*. **336:**823–827, 1997.
68. Holtzman, N. Effect of education on physicians' knowledge of laboratory screening tests. *J. Clin. Ethics* **2:**241–243, 1992.
69. Hunter, A., Wright, P., Cappelli, M., Kasaboski, A., and Surh, L. Physician knowledge and attitudes towards molecular genetic (DNA) testing of their patients. *Clin. Genet*. **53:**447–455, 1998.
70. Sandhouse, L. M., Singer, M. E., Dawson, N. V., and Wiesner, G. L. Reporting BRCA test results to primary care physicians. *Gen. Med*. **3:**327–334, 2001.
71. Escher, M. and Sappino, A. P. Primary care physicians' knowledge and attitudes towards genetic testing for breast–ovarian cancer predisposition. *Ann. Oncol*. **11:**1131–1135, 2000.
72. Geller, G. and Holtzman, N. A. A qualitative assessment of primary care physicians perceptions about the ethical and social implications of offering genetic testing. *Qual.. Health Res*. **5:**97–116, 1995.
73. Press, N. A., Yasui, Y., Reynolds, S., Durfy, S. J., and Burke, W. Women's interest in genetic testing for breast cancer susceptibility may be based on unrealistic expectations. *Am. J. Med. Genet*. **99:**99–110, 2001.
74. Marteau, T. Towards informed decisions about prenatal testing: a review. *Prenat. Diagn*. **15:**1215–1226, 1995.
75. Marteau, T. M. and Dormandy, E. Facilitating informed choice in prenatal testing: how well are we doing? *Am. J. Med. Genet*. **106:** 185–190, 2001.
76. Senior, V., Marteau, T. M., and Peters, T. J. Will genetic testing for predisposition for disease result in fatalism? A qualitative study of parents' responses to neonatal screening for familial hypercholesterolemia. *Soc. Sci. Med*. **48:**1857–1860, 1999.
77. Michie, S., Smith, J., Senior, V., and Marteau, T. M. Understanding why negative genetic test results sometimes fail to reassure. *Am. J. Med. Genet*. **119A:**340–347, 2003.
78. Roche, M., Skinner, D., Kuczynski, K., and Shaffer, R. Parents' use of the internet as a source of genetic information. *J. Genet. Counsel*. **11:**506a, 2002.
79. Rosenthal, E., Biesiecker, L., and Biesecker, B. Parental attitudes toward a diagnosis in children with unidentified multiple congenital anomaly syndromes. *Am. J. Med. Genet*. **103:**106–114, 2001.
80. Magnuson, M. and Hergils, L. Late diagnosis of congenital hearing impairment in children: the parents' experiences and opinions. *Patient Educ. Counsel.* **41:**285–294, 2000.

81. Genetic Evaluation of Congenital Hearing Loss Expert Panel ACMG Statement: genetics evaluation guidelines for the etiologic diagnosis of congenital hearing loss. *Genet. Med*. **4:**162–171, 2002.
82. Baker, D., Schuette, J., and Uhlmann, W. eds., *Guide to Genetic Counseling.* Wiley–Liss, New York, 1998.
83. Schnieder, K. *Counseling About Cancer*. Wiley-Liss, New York, 2002.
84. American Society of Clinical Oncology Policy Statement Update: Genetic Testing for Cancer Susceptibility. *J. Clin Oncol*. **21:**2397–2406, 2003.
85. Marymee, K., Dolan, C., Pagon, R., Bennett, R., Coe, S., and Fisher, N. Development of the critical elements of genetic evaluation and genetic counseling for genetic professionals and perinatologists in Washington State. *J. Genet. Counsel*. **7:**133–165, 1998.
86. Bennett, R., Petterson, B., Niendorf, K., and Anderson, R. Standard recommendations (guidelines) for genetic counseling practices: a process of the National Society of Genetic Counselors. *J. Genet. Counsel*. **12:**287–295, 2003.
87. Bennett, R. *The Practical Guide to the Genetic Family History* Wiley-Liss, New York, 1999.
88. Kessler, S. The process of communication, decision-making and coping in genetic counseling, in *Genetic Counseling: Psychological Dimensions,* Kessler, S., ed., Academic, New York, pp. 35–52, 1979.
89. Weil, J. *Psychosocial Genetic Counseling*. Oxford University Press, New York, 2000.
90. Shannon, K., Lubratovich, M., Finkelstein, D., Smith, D., Powell, S., and Seiden, M. Model-based predictions of BRCA1/2 mutation status in breast carcinoma patients treated at an academic medical center. *Cancer* **94:**305–313, 2002.
91. Meschede, D., Albersmann, S., and Horst, J. The practical importance of pedigree analysis in women considering invasive prenatal diagnosis for advanced maternal age or abnormal serum screening tests. *Prenat. Diagn*. **20:**865–869, 2000.
92. Henneman, L., Bramsen, I., Van Os, T., et al. Attitudes towards reproductive issues and carrier testing among adult patients and parents of children with cystic fibrosis (CF) *Prenat. Diagn*. **21:**1–9, 2001.
93. Broadstock, M., Michie, S., and Marteau, T. Psychological consequences of predictive genetic testing: a systematic review. *Eur. J. Hum. Genet.* **8:**731–738, 2000.
94. Hughes, C., Lerman, C., Schwartz, M., et al. All in the family: evaluation of the process and content of sisters' communication about BRCA1 and BRCA2 genetic test results. *Am. J. Med. Genet*. **107:**143–150, 2002.
95. van Rijn, M., de Vries, B., Tibben, A., van den Ouweland, A., Halley, D., and Niermeijer, M. DNA testing for fragile X syndrome: Implications for parents and family. *J. Med. Genet*. **34:**907–911, 1997.
96. Ormond, K. Effect of family history on disclosure patterns of CF carrier status. *Am. J. Med. Genet*. **119C:**70–77, 2003.
97. Fanos, J. My crooked vision: the well sib views ataxia-telangiectasia. *Am. J. Med. Genet*. **87:**420–425, 1999.
98. Fanos, J., Davis, J., and Puck, J. Sib understanding of genetics and attitudes toward carrier testing for X-linked severe combined immunodeficiency. *Am. J. Med. Genet*. **98:**46–56, 2001.
99. Hunt, K., Emslie, C., and Watt, G. Lay constructions of a family history of heart disease: a potential for misunderstanding in the clinical encounter? *Lancet* **357:**1168–1171, 2001.
100. Parsons, E. P. and Atkngson, P. A. Lay constructions of genetic risk. *Sociol. Health Illness* **14:**437–455. 1992.
101. Weil, J. Multicultural education and genetic counseling. *Clin. Genet*. **59:**143–149, 2001.
102. Browner, C., Preloran, M., Casado, M., Bass, H., and Walker, A. Genetic counseling gone awry: miscommunication between prenatal genetic service providers and Mexican-origin clients. *Soc. Sci. Med*. **56:**1933–1946, 2003.
103. Collins, V., Halliday, J., Kahler, S., and Williamson, R. Parents' experiences with genetic counseling after the birth of a baby with a genetic disorder: an exploratory study. *J. Genet. Counsel*. **10:**53–72, 2001.
104. Anderson, V. E. Obituary: Sheldon Reed, Ph.D.: genetic counseling, behavioral genetics. *Am. J. Hum. Genet*. **73:**1–4, 2003.
105. Reed, S. *Counseling in Medical Genetics*. Saunders, Philadelphia, PA, 1957.
106. Resta, R. Sheldon Clark Reed, Ph.D. in memorium. *J. Genet. Counsel*. **12:**283–285, 2003.
107. Sikkens, E., de Walle, H., Reefhuis, J., Tintelen, J., and van Essen, A. Referral for genetic counseling after the birth of a child with a congenital anomaly in the northern Netherlands. *Am. J. Med. Genet*. **112:**133–137, 2002.
108. Geller, G. Tambour, E. S. Chase, G. A., Hofman, K. J., Faden, R. R., and Holtzman, N. A. Incorporation of genetics in primary care practice: will physicians do the counseling and will they be directive? *Arch. Fam. Med*. **2:**1119–1125, 1993.
109. Collins, F. Preparing health professionals for the genetic revolution. *J. Am. Med. Assn.* **278:**1285–1286, 1997.
110. Facher, J. and Robin, N. Genetic counseling in primary care. What questions are patients likely to ask and how should they be answered? *Postgrad. Med*. **107:**59–66, 2000.
111. Yoon, P. W., Scheuner, M. T., Peterson-Oehlke, K. L., Gwinn, M., Faucett, A., and Khoury, M. J. Can family history be used as a tool for public health and preventive medicine? *Genet. Med*. **4:**304–310, 2002.
112. Fraser, F. Genetic counseling. *Am. J. Hum. Genet*. **26:**636–661, 1974.
113. Biesecker, B. and Marteau, T. The future of genetic counseling: an international perspective. *Nat. Genet*. **22:**133–137, 1999.
114. Walker, A. The practice of genetic counseling, in *A Guide to Genetic Counseling,* Baker, D., Shuette, J., and Uhlmann, W., eds., Wiley–Liss, New York, 1999.
115. National Society of Genetic Counselors. National Society of Genetic Counselors Code of Ethics. *J. Genet Counsel*. **1:**41–44, 1992.
116. Kessler, S. Introduction., in *Genetic Counseling: Psychological Dimensions,* Kessler, S., ed., Academic, New York, 1979, pp. 1–15.
117. Fine, B. The evolution of non-directivenss in genetic counseling and implications of the Human Geneome Project, in *Prescribing Our Future: Ethical Challenges in Genetic Counseling,* Bartels, D., LeRoy, B., and Caplan, A., eds., Aldine de Gruyter, New York, 1993, pp. 101–117.
118. Press, N. A. and Bowner, C. H. 'Collective fictions': similarities in reasons for accepting maternal serum alpha-fetoprotein screening among women of diverse ethnic and social class backgrounds. *Fetal Diagn. Ther*. **8:**97–106, 1993.
119. Rapp, R. Communicating about chromosomes: patient, providers, and cultural assumptions. *J. Am. Med. Women's Assoc*. **52:**28–32, 1997.
120. Wertz, D. C. and Fletcher, J. C. Ethics and medical genetics in the United States: a national survey. *Am. J. Hum. Genet*. **29:**815–827, 1988.
121. Marteau, T., Plenicar, M., and Kidd, J. Obstetricians presenting amniocentesis to pregnant women: practice observed. *J. Reprod. Infant. Psychol*. **11:**3–10, 1993.
122. McCarthy-Veach, P., Bartels, D., and LeRoy, B. Ethical and professional challenges posed by patients with genetic concerns: a report of focus group discussions with genetic counselors, physicians, and nurses. *J. Genet. Counsel*. **10:**97–119, 2001.
123. Harris, R. Concerted action on genetic services in Europe: a comparative study of 31 countries. *Eur. J. Hum. Genet*. **5:**1–220, 1997.
124. Fiddler, M., Fine, B., Baker, D., and the ABGC Consensus Development Consortium. Practice-based competencies for accreditation of and training in graduate programs in genetic counseling. *J. Genet. Counsel*. **5:**105–112, 1996.
125. Begleiter, M. Training for genetic counselors. *Nat. Rev. Genet*. **3:**557–561, 2002.
126. Guttmacher, A., Jenkins, J., and Uhlmann, W. Genomic medicine: who will practice it? A call to open arms. *Am. J. Med. Genet*. **106:**216–222, 2001.
127. Bowen, D., Ludman, E., Press, N., Vu, T., and Burke, W. Achieving utility with family history—colorectal cancer risk. *Am. J. Prev. Med*. **24:**177–182, 2003.
128. Harrison, T. A., Hindorff, L. A., Kim, H., et al. Family history of diabetes as a potential public health tool. *Am. J. Prev. Med*. **24:**152–159, 2003.

129. Henderson, L. and Scheuner, M. Suboptimal family history collection: an obstacle to genetic risk assessment and prevention. *Am. J. Hum. Genet*. **63:**A340, 1998.
130. Leffett, V., MacKay, J., Marteau, T., and Yates, J. The psychological impact of a cancer family history questionnaire completed in general practice. *J. Med. Genet*. **37:**470–472, 2000.
131. Acheson, L., Wiesner, G., Zyzanski, S., Goodwin, M., and Stange, K. Family history-taking in community family practice: Implications for genetic screening. *Genet. Med*. **2:**180–185, 2000.
132. Austin, M., Petser, P., and Khoury, M. The interface of genetics and public health: research and educational challenges. *Ann. Rev. Public Health* **21:**81–99, 2000.
133. Greendale, K. and Pyeritz, R. Empowering primary care health professionals in medical genetics: how soon? How fast? How far? *Am. J. Med. Genet*. **106:**223–232, 2001.
134. Pencarinha, D., Bell, N. K., Edwards, J. G., and Best, R. G. Ethical issues in genetic counseling: a comparison of MS counselor and medical geneticist perspectives. *J. Genet. Counsel*. **1:**19–30, 1992.
135. Lucassen, A. and Houlston, R. Clinical geneticists' attitudes and practice towards testing for breast cancer susceptibility genes. *J. Med. Genet*. **37:**157–160, 2000.
136. Resta, R. Eugenics and nondirectiveness in genetic counseling. *J. Genet. Counsel.* **6:**255–258, 1997.
137. Biesecker, B. Goals of genetic counseling. *Clin. Genet*. **69:**323–330, 2001.
138. Biesecker, B. and Peters, K. Process studies in genetic counseling: peering into the black box. *Am. J. Med. Genet*. **106:**191–198, 2001.
139. Clarke, A. Is nondirective genetic counselling possible? *Lancet* **338:**998–1001, 1991.
140. Michie, S., Bron, F., Bobrow, M., and Marteau, T. M. Nondirectiveness in genetic counseling: an empirical study. *Am. J. Hum. Genet*. **60:**40–47, 1997.
141. Bartels, D., LeRoy, B., McCarthy, P., and Caplan, A. Nondirectiveness in genetic counseling: a survey of practitioners. *Am. J. Med. Genet*. **72:**172–179, 1997.
142. Shiloh, S. and Saxe, L. Perception of risk in genetic counseling. *Psychol. Health* **3:**45–61, 1989.
143. Weil, J. Psychosocial genetic counseling in the post-nondirective era: a point of view. *J. Genet. Counsel*. **12:**199–211, 2003.
144. Wertz, D. Geneticists' ethical views: a survey in 37 nations, in European Society of Human Genetics, Berlin, May 24–27, 1995.
145. Farrell, M. H., Certain, L. K., and Farrell, P. M. Genetic counseling and risk communication services of newborn screening programs. *Arch. Pediatr. Adolesc. Med*. **155:**120–126, 2001.
146. Lippman-Hand, A. and Fraser, F. Genetic counseling: the postcounseling period. II. Making reproductive choices. *Am. J. Med. Genet*. **4:**73–87, 1979.
147. Biesecker, B., Benkendorf, J., Kerner, J., Gomez-Caminero, A., Hughes, C., and Reed, M. Controlled trial of pretest education approaches to enhance informed decision-making for BRCA1 gene testing. *J. Natl. Cancer Inst*. **89:**148–157, 1997.
148. Loeben, G., Marteau, T., and Wilfond, B. Mixed messages: presentation of information in cystic fibrosis screening pamphlets. *Am. J. Hum. Genet*. **63:**1181–1189, 1998.
149. Geller, G., Botkin, J. R., Green, M. J., et al. Genetic testing for susceptibility to adult-onset cancer: the process and content of informed consent. *J. Am. Med. Assoc.* **277:**1467–1474, 1997.
150. Watson, M., Duvivier, V., Walsh, M. W., et al. Family history of breast cancer: what do women understand and recall about their genetic risk? *J. Med. Genet*. **35:**731–738, 1998.
151. Green M. J., Biesecker B. B., McInerney, A. M., Mauger, D., and Fost, N. An interactive computer program can effectively educate patients about genetic testing for breast cancer susceptibility. *Am. J Med. Genet*. **103:**16–23, 2001.
152. Westman, J., Hampel, H., and Bradley, T. Efficacy of a touchscreen computer based family cancer history questionnaire and subsequent cancer risk assessment. *J. Med. Genet*. **37:**354–360, 2000.
153. Emery, A. E. H., Raeburn, J. A., Skinner, R., Holloway, S., and Lewis, M. Prospective study of genetic counseling. *Br. Med. J*. **1:**1253–1256, 1979.
154. Ekow, E. E., Kim, J. O., and Gosselink, C. A. Parental perception of the burden of genetic disease. *Am. J. Med. Genet*. **28:**955–963, 1987.
155. Michie, S. and Marteau, T. The choice to have a disabled child. *Am. J. Hum. Genet.* **63:**1175–1180, 1999.
156. Marteau, T. and Johnson, M. Determinants of beliefs about illness: a study of parents of children with diabetes, asthma, epilepsy, and no chronic illness. *J. Psychosom. Res*. **30:**1175–1180, 1986.
157. Croyle, R. T. and Lerman, C. Interest in genetic testing and colon cancer susceptibility: cognitive and emotional correlates. *Prev. Med*. **22:**284–292, 1993.
158. Mischler, E., Wilfond, B., Fost, N., et al. Cystic fibrosis newborn screening: impact on reproductive behavior and implications for genetic counseling. *Pediatrics* **102:**44–52, 1998.
159. Evers-Kiebooms, G. and van den Berghe, H. Impact of genetic counseling; a review of published follow up studies. *Clin. Genet*. **15:**465–474, 1979.
160. Wertz, D. C. and Sorenson, J. R. Genetic counseling and reproductive uncertainty. *Am. J. Med. Genet*. **18:**79–88, 1984.
161. Frets, P. G., Duivenvoorden, H. J., Verhage, F., Niemeijer, M. F., van de Berge, S. M., and Galjaard, H. Factors influencing the reproductive decision after genetic counseling. *Am. J. Med. Genet*. **35:**496–502, 1990.
162. Lippman-Hand, A. and Fraser, F. Genetic counseling; the postcounseling period. I. Parents' perceptions of uncertainty. *Am. J. Med. Genet*. **4:**51–71, 1979.
163. Oetting, L. and Steele, M. A controlled retrospective follow-up study of the impact of genetic counselling on parental reproduction following the birth of a Down syndrome child. *Clin. Genet*. **21:**7–13, 1982.
164. Somer, M., Mustonene, H., and Norio, R. Evaluation of genetic counseling: recall of information, post-counseling reproduction, and attitudes of the counselees. *Clin. Genet*. **34:**352–365, 1988.
165. Swerts, A. Impact of genetic counseling and prenatal diagnosis for Down syndrome and neural tube defects, *Birth Defects: Original Article Series*, March of Dimes Birth Defects Foundation, **23:**61–83, 1987.
166. Nossal, G. and Ross, C. *Reshaping Life: Key Issues in Genetic Engineering,* 3rd ed., Cambridge University Press, Cambridge, 2002.
167. Beeson, D. Nuance, complexity, and content: qualitative methods in genetic counseling research, *J. Genet. Counsel*. **6:**21–43, 1997.
168. DudokdeWit, A. C., Tibben, A., Duivenvoorden, H., Niermeijer, M. F., Passchier, J. R., and Trijsburg, R. Distress in individuals facing predictive DNA testing for autosomal dominant late-onset disorders: comparing questionnaire results with in-depth interviews. *Am. J. Med. Genet*. **75:**62–74, 1998.
169. Reyna, V., Lloyd, F., and Whalen, P. Genetic testing and medical decision making. *Arch. Intern. Med*. **161:**2406–2408, 2001.
170. Clarke, A., Parsons, E., and Williams, A. Outcomes and process in genetic counseling. *Clin. Genet*. **50:**462–469, 1996.
171. Shiloh, S., Avdor, O., and Goodman, R. M. Satisfaction with genetic counseling: dimensions and measurement. *Am. J. Med. Genet*. **37:**522–529, 1990.
172. Michie, S., McDonald, V., and Marteau, T. M. Genetic counseling: information given, recall, and satisfaction. *Patient Educ. Counsel*. **32:**101–106, 1997.
173. Michie, S., Marteau, T., and Bobrow, M. Genetic counseling: the psychological impact of meeting patients' expectations. *J. Med. Genet*. **34:**237–241, 1997.
174. Kash, K. M., Holland, J. C., Halper, M. S., and Miller, D. G. Psychological distress and surveillance behaviors of women with a family history of breast cancer. *J. Natl. Cancer Inst*. **84:**24–30, 1992.
175. Lerman, C., Hughes, C., Croyle, R. T., et al. Prophylactic surgery decisions and surveillance practices one year following BRCA 1 and 2 testing. *Prevent Med*. **31:**78–80, 2000.
176. Armstrong, D., Michie, S., and Marteau, T. Revealed identity: A study of the process of genetic counseling. *Soc. Sci. Med*. **47:**1653–1658, 1998.
177. Hallowell, N., Murton, F., Statham, H., Green, J. M., and Richards, M. P. M. Women's need for information before attending

genetic counseling for familial breast or ovarian cancer: a questionnaire, interview, and observational study. *Br. Med. J*. **314:**281–283, 1997.
178. Bernhardt, B. A., Biesecker, B. B., and Mastromarino, C. Goals, benefits, and outcomes of genetic counseling: client and genetic counselor assessment. *Am. J. Med. Genet*. **94:**189–197, 2000.
179. Wertz, D. and Gregg, R. Genetic services in a social, ethical, and policy context: a collaboration between consumers and providers. *J. Med. Ethics* **26:**261–265, 2000.
180. Lerman, C., Narod, S., Schulman, K., et al. BRCA1 testing in families with hereditary breast–ovarian cancer: a prospective study of patient decision making and outcomes. *J. Am. Med. Assoc*. **275:**1885–1892, 1996.
181. Baker, D., Eash, T., Schuette, J., and Uhlmann, W. Guidelines for writing letters to patients. *J Genet. Counsel*. **11:**399–416, 2002.
182. McKinnon, W. C., Baty, B. J., Bennett, R. L., et al. Predisposition genetic testing for late-onset disorders in adults—a position paper of the National Society of Genetic Counselors. *J. Am. Med. Assoc*. **278:**1217–1220, 1997.
183. Statement of American Society of Human Genetics on Genetic Testing for Breast and Ovarian Cancer Predisposition *Am. J. Hum. Genet*. **55:**i–iv, 1994.
184. Joint Commission on Infant Hearing. Year 2000 position statement: principles and guidelines for early hearing detection and intervention programs. *Pediatrics* **106:**798–817, 2000.
185. International Huntington Association and the World Federation of Neurology Research Group on Huntington's Chorea. Guidelines for the molecular genetics predictive test in Huntington's disease. *J. Med. Genet*. **31:**555–559, 1994.
186. Green, M. J. and Botkin, J. R. Genetic "exceptionalism" in medicine: clarifying the differences between genetic and nongenetic tests. *Ann. Intern. Med*. **138:**571–575, 2003.
187. McGovern, M., Benach, M., and Zinberg, R. Interaction of genetic counselors with molecular genetic testing laboratories: implications for non-geneticist health care providers. *Am. J. Med. Genet*. **119A:**297–301, 2003.
188. Holtzman, N., Hilgartner, S., Brierley-Bowers, K., and Tambour, E. Survey of biotechnology companies (BTC's) and molecular and cytogenetics laboratories (LABs) on genetic testing. *Am J. Hum. Genet*. **57:**A3, 1995.
189. Amos, J. and Gold, B. Testing environment for single-gene disorders in US reference laboratories. *Hum. Mutat*. **12:**293–300, 1998.
190. Andersson, H., Krousel-Wood, M., Jackson, K., Rice, J., and Lubin, I. Medical genetic test reporting for cystic fibrosis (Delta F508) and factor V Leiden in North American laboratories. *Genet. Med*. **4:**324–327, 2002.
191. Roche, M., Aylsworth, A., Lachicotte, W., Bailey, D., and Skinner, D. Parental reactions to genetic testing results. *J. Genet. Counsel*. **10:**481A, 2001.
192. Christian, S., Kieffer, S., and Leonard, N. Medical genetics and patient use of the Internet. *Clin. Genet*. **60:**232–236, 2001.
193. Taylor, M., Alman, A., and Manchester, D. Use of the Internet by patients and their families to obtain genetics-related information. *Mayo Clin. Proc*. **76:**772–776, 2001.
194. Culver, J., Burke, W., Yasui, Y., Durfy, S., and Press, N. Participation in breast cancer genetic counseling: the influence of Educational level, ethnic background, and risk perception. *J. Genet. Counsel*. **10:**215–231, 2001.
195. Halliday, J., Griffin, O., Bankier, A., Rose, C., and Riley, M. Use of record linkage between a statewide genetics service and a birth defects congenital malformations register to determine use of genetic counseling services. *Am. J. Med. Genet*. **72:**3–10, 1997.
196. Bernhardt, B., Geller, G., Doksum, T., Larson, S., Roter, D., and Holtzman, N. Prenatal genetic testing: content of discussions between obstetric providers and pregnant women. *Obstet. Gynecol*. **91:**648–655, 1998.
197. Collins, V., Halliday, J., and Williamson, R. What predicts the use of genetic counseling services after the birth of a child with Down syndrome? *J. Genet. Counsel*. **12:**43–60, 2003.
198. Hayflick, S. J., Eiff, M. P., Carpenter, L., and Steinberger, J. Primary care physicians' utilization and perceptions of genetic services. *Genet. Med*. **1:**13–21, 1998.
199. Fanos, J. Developmental tasks of childhood and adolescence: implications for genetic testing. *Am. J. Med. Genet*. **71:**22–28, 1997.
200. Wagener, J., Wagener, S., Farrell, P., and Corey, M. A debate on why my state (province) should or should not conduct newborn screening for cystic fibrosis. *Pediatr. Pulmonol.* **32:**385–396, 2001.
201. Bradley, D., Parsons, E., and Clarke, A. Experience with screening newborns for Duchenne muscular dystrophy in Wales. *Br. Med. J*. **306:**357–360, 1993.
202. Thelin, T., McNeil, T., Aspegren-Jansson, E., and Sveger, T. Psychological consequences of neonatal screening for alpha 1 antitrypsin deficiency. *Acta Pardiatr. Scand*. **74:**841–847, 1985.
203. Burke W. Genomic medicine: genetic testing. *N. Engl. J. Med*. **347:**1867–1875, 2002.
204. Pesso, R., Berkenstadt, H., Cuckle, H., et al. Screening for fragile X syndrome in women of reproductive age. *Prenat. Diagn*. **20:**611–614, 2000.
205. Sherman, S. L. Expansion of the fragile X CGG repeat in females with premutation or intermediate alleles. *Am. J. Hum. Genet*. **72:**454–464, 2003.
206. American Academy of Pediatrics, Committee on Bioethics. Ethical issues with genetic testing in Pediatrics. *Pediatrics* **107:**1451–1455, 2001.
207. Wingrove, K., Norris, J., Barton, P., and Hagerman, R. Experiences and attitudes concerning genetic testing and insurance in a Colorado population: a survey of families diagnosed with fragile X syndrome. *Am. J. Med. Genet*. **64:**378–381, 1996.
208. McConkie-Rosell, A., Spiridigliozzi, G., Sullivan, J., Dawson, D., and Lachiewicz, A. Carrier testing in fragile X syndrome: when to tell and test. *Am. J. Med. Genet.* **110:**36–44, 2002.
209. McConkie-Rosell, A., Spiridigliozzi, G., et al. Parental attitudes regarding carrier testing in children at risk for fragile X syndrome. *Am. J. Med. Genet*. **82:**206–211, 1999.
210. Werz, D. C. and Reilly, P. R. Laboratory policies and practices for the genetic testing of children: a survey of the helix network. *Am. J. Hum. Genet*. **61:**1163–1168, 1997.
211. Steele, K. and Cross, C. A genetic approach to understanding auditory function. *Nat. Genet*. **27:**143–149, 2001.
212. Resendes, B., Williamson, R., and Morton, C. At the speed of sound. *Am. J. Hum. Genet*. **69:**923–935, 2001.
213. Willems, P. Genetic causes of hearing loss. *N. Engl. J. Med*. **342:**1101–1109, 2000.
214. Van Camp, G., Williams, P., and Smith, R. Nonsyndromic hearing loss: unparalleled heterogeneity. *Am. J. Hum. Genet*. **60:**758–764, 1997.
215. Morton, N. Genetic epidemiology of hearing loss. *Ann. NY Acad. Sci*. **630:**16–31, 1991.
216. Hone, S. and Smith, R. Medical evaluation of pediatric hearing loss: laboratory, radiographic and genetic testing. *Otolaryngol. Clin. North Am*. 35:751–764, 2002.
217. Green, G., Scott, D., McDonald, J., et al. Performance of cochlear implant recipients with GJB2-related deafness. *Am. J. Med. Genet.* **109:**167–170, 2002.
218. Cohn, E. and Kelley, P. Clinical phenotype and mutations in connexin 26 (DFNB1/GJB2), the most common cause of childhood hearing loss. *Am. J. Med. Genet.* **89:**130–136, 1999.
219. Greene, G., Scott, D., McDonald, J., Woodworth, G., Sheffield, V., and Smith, R. Carrier rates in the Midwestern United States for GJB2 mutations causing inherited deafness. *J. Am. Med. Assoc*. **281:**2211–2216, 1999.
220. Moeller, M. Early intervention and language development in children who are deaf and hard of hearing. *Pediatrics* **106:**e43, 2000.
221. Yoshinaga-Itano, C., Sedey, A., Coulter, D., and Mehl, A. Language of early and later-identified children with hearing loss. *Pediatrics* **102:**1161–1171, 1998.
222. Brunger, J., Murray, G., O'Riordan, M., Matthews, A., Smith, R., and Robin, N. Parental attitudes toward genetic testing for pediatric deafness. *Am. J. Hum. Genet.* **67:**1621–1625, 2000.

223. Middleton, A., Hewison, J., and Mueller, R. Attitudes of deaf adults toward genetic testing for hereditary deafness. *Am. J. Hum. Genet*. **63:**1175–1180, 1998.
224. Antoniadi, T., Pampanos, A., and Petersen, M. Prenatal diagnosis of prelingual deafness: carrier testing and prenatal diagnosis of the common GJB2 35delG mutation. *Prenat. Diagn*. **21:**10–13, 2001.
225. Ryan, M., Miedzybrodzka, Z., Fraser, L., and Hall, M. Genetic information but not termination: pregnant women's attitudes and willingness to pay for carrier screening for deafness genes. *J. Med. Genet*. **40:**e80, 2003.
226. Martinez, A., Linden, J., Schimmenti, L., and Palmer, C. Attitudes of the broader hearing, deaf, and hard of hearing community toward genetic testing for deafness. *Genet. Med*. **5:**106–112, 2003.
227. Dagan, O., Hochner, H., Levi, H., Raas-Rothschild, A., and Sagi, M. Genetic testing for hearing loss: different motivations for the same outcome. *Am. J. Med. Genet*. **113:**137–143, 2002.
228. Barrette, J. and Marden, C. Attitudes of families to some aspects of Huntington's chorea. *Psychol. Med*. **9:**327–336, 1979.
229. Bundy, S. Few psychological consequences of presymptomatic testing for Huntington disease. *Lancet* **349:**4, 1997.
230. Cannella, M., Simonelli, M., D'Alessio, C., Pierelli, F., Ruggieri, S., and Squitieri, F. Presymtomatic testing in Huntington disease and dominant ataxias. *Neurol. Sci*. **22:**55–56, 2001.
231. Codori, A. M. and Brandt, J. Psychological costs and benefits of predictive testing for Huntington disease. *Am. J. Med. Genet*. **54:**174–184, 1994.
232. Cox, S. and McKellin, W. 'There's this thing in our family': Predictive testing and the construction of risk for Huntington disease. *Sociol. Health Illness* **21:**622–646, 1999.
233. Harper, P. S., Ball, D. M., and Hayden, M. R. Presymptomatic testing for Huntington's Disease: a worldwide survey. *J. Med. Genet*. **30:**1020–1022, 1993.
234. Petersen, G. and Bresinger, J. Genetic testing and counseling in familial adenomatous polyposis. *Oncology* **10:**89–94, 1996.
235. Jedlicka-Kohler, I., Gotz, M., and Eichler, I. Utilization of prenatal diagnosis for cystic fibrosis over the past seven years. *Pediatrics* **94:**13–16, 1994.
236. Marteau, T. M. and Croyle, R. T. Psychological responses to genetic testing. *Br. Med. J*. **316:**693–696, 1998.
237. Gwyn, K., Vernon, S., and Conoley, P. Intention to pursue genetic testing for breast cancer among women due for screening mammography. *Cancer Epidemiol. Biomarkeis Prev*. **12:**96–102, 2003.
238. Tambour, E. S., Bernhardt, B. A., and Chase, G. A. Offering cystic fibrosis carrier screening to an HMO population: factors associated with utilization. *Am. J. Hum. Genet.* **55:**626–637, 1994.
239. Cappelli, M., Suhr, L., Humphreys, L., et al. Psychological and social determinants of women's decisions to undergo genetic counseling and testing for breast cancer. *Clin. Genet*. **55:**419–430, 1999.
240. Lenoir, G. and Lynch, H. BRCA1 testing in families with hereditary breast and ovarian cancer: a prospective study of patient decision-making and outcomes. *J. Am. Med. Assoc*. **275:**1885–1892, 1996.
241. Lerman, C., Peters, J., Ades, T., et al. Genetic counseling issues. *Cancer* **80:**628–629, 1997.
242. Lerman, C., Daly, M., Masny, A., and Balshem, A. Attitudes about genetic testing for breast and ovarian cancer susceptibility. *J. Clin. Oncol*. **12:**843–850, 1994.
243. Geller, G., Bernhardt, B., Doksum, T., Helzlsouer, K., Wilcox, P., and Holtzman, N. Decision-making about breast cancer susceptibility testing: how similar are the attitudes of physicians, nurse practitioners, and at-risk women? *J. Clin. Oncol*. **16:**1–10, 1998.
244. Glanz, K., Grove, J., Lerman, C., Gotay, C., and LeMarchand, L. Correlates of intention to obtain genetic counseling and colorectal cancer gene testing among at-risk relatives from three ethnic groups. *Cancer Epidemiol. Biomarkers Prev*. **8:**329–336, 1999.
245. Biesecker, B. B., Ishibe, N., Hadley, D. W., et al. Psychosocial factors predicting BRCA1/BRCA2 testing decisions in members of hereditary breast and ovarian cancer families. *Am. J. Med. Genet.* **93:**257–263, 2000.
246. Vernon, S., Gritz, E., Petersen, S., et al. Intention to learn results of genetic testing for hereditary colon cancer. *Cancer Epidemiol. Biomarkers. Prev*. **8:**353–360, 1999.
247. Sujansky, E., Kreutzer, S., Johnson, A., Lezotte, D., Schrier, R., and Gabow, P. Attitudes of at-risk and affected individuals regarding presymptomatic testing for autosomal dominant polycystic kindey disease. *Am. J. Med. Genet*. **34:**510–515, 1990.
248. Duisterhof, M., Trijsburg, R., Niermeijer, M., Roos, R., and Tibben, A. Psychological studies in Huntington's disease: making up the balance. *J. Med. Genet.* **38:**852–861, 2001.
249. Bekker, H., Denniss, G., Modell, M., Bobrow, M., and Marteau, T. M. The impact of population-based screening for carriers of cystic fibrosis. *J. Med. Genet*. **31:**364–368, 1994.
250. Sorenson, J., Cheuvront, B., DeVellis, B., et al. Acceptance of home and clinic-based cystic fibrosis carrier education and testing by first, second, and third degree relatives of cystic fibrosis patients. *Am. J. Med. Genet*. **70:**121–129, 1997.
251. Bernhardt, B., Tambour, E., Fraser, G., Wissow, L., and Geller, G. Parents' and children's attitudes toward the enrollment of minors in genetic susceptibility research: implications for informed consent. *Am. J. Med. Genet*. **116A:**315–323, 2003.
252. Durfy, S., Bowen, D., McTierman, A., Sporleder, J., and Burke, W. Attitudes and interest in genetic testing for breast and ovarian cancer susceptibility in diverse groups of women in Western Washington. *Cancer Epidemiol Biomarkers Prev*. **8:**369–375, 1999.
253. Lafayette, D., Abuelo, D., Passero, M., and Tantravahi, U. Attitudes towards cystic fibrosis carrier and prenatal testing and utilization of carrier testing among relatives of individuals with cystic fibrosis. *J. Genet. Counsel*. **8:**17–36, 1999.
254. Bottorff, J. L., Ratner, P. A., Balneaves, L. G., et al. Women's interest in genetic testing for breast cancer risk: the influence of sociodemographics and knowledge. *Cancer Epidemiol Biomarkers Prev*. **11:** 89–95, 2002.
255. Meiser, B., Gleeson, M., and Tucker, K. Psychological impact of genetic testing for adult-onset disorders. *Med. J. Austr.* **172:** 126–129, 2000.
256. Codori, A., Hanson, R., and Brandt, J. Self-selection in predictive testing for Huntington disease. *Am. J. Med. Genet*. **54:**167–173, 1994.
257. Grosfeld, F. J. M., Lips, C. J. M., Beemer, F. A., et al. Distress in MEN 2 family members and partners prior to DNA test disclosure. *Am. J. Med. Genet*. **91:**1–7, 2000.
258. Butow, P. N., Lobb, E. A., Meiser, B., Barratt, A., and Tucker, K. M. Psychological outcomes and risk perception after genetic testing and counseling in breast cancer: a systematic review. *Med. J. Aust*. **178:**77–81, 2003.
259. Collins, V., Halliday, J., Warren, R., and Williamson, R. Cancer worries, risk perceptions and associations with interest in DNA testing and clinic satisfaction in a familial colorectal cancer clinic. *Clin. Genet*. **58:**460–468, 2000.
260. Kenen, R. and Schmidt, R. Stigmatization of carrier status: social implications of heterozygote genetic screening programs. *Am. J. Public Health*. **68:**116–120, 1978.
261. Craufurd, D., Kerzin-Starrar, L., Dodge, A., and Harris, R. Uptake of presymptomatic predictive testing for Huntington disease. *Lancet* **21:**603–605, 1989.
262. Roos, R., Vedter-van der Vlis, M., Hermans, J., et al. Age at onset in Huntington's disease: effect of line of inheritance and patient's sex. *J. Med. Genet*. **28:**515–519, 1991.
263. Brandt, J., Quaid, K., Folstein, S., et al. Presymptomatic diagnosis of delayed-onset disease with linked DNA markers: the experience in Huntington's disease. *J. Am. Med. Assoc*. **261:**3108–3114, 1989.
264. DudokdeWit, A. C., Tibben, A., Duivenvoorden, H. J., Niermeijer, M. F., and Passchier, J. Predicting adaptation to presymptomatic DNA testing for late onset disorders: who will experience distress? *J. Med. Genet*. **35:**745–754, 1998.
265. Quaid, K. Presymptomatic testing for Huntington Disease: recommendations for counseling. *J. Genet. Counsel*. **1:**277–302, 1992.
266. Evers-Kiebooms, G., Swerts, A., Cassiman, J. J., and Van Den Berghe, H. The motivation of at-risk individuals and their partners in deciding for or against predictive testing for Huntington Disease. *Clin. Genet*. **35:**29–40, 1989.

267. Falush, D., Almqvist, E., Brinkmann, R., Iwasa, Y., and Hayden, M. Measurement of mutational flow implies both a high new mutation rate for Huntington disease and substantial under ascertainment of late-onset cases. *Am. J. Hum. Genet*. **68:**373–385, 2000.
268. The American College of Medical Genetics/American Society of Human Genetics Huntington Disease Genetic Testing Working Group. ACMG/ASHG Statement: laboratory guidelines for Huntington disease genetic testing. *Am. J. Hum. Genet.* **62:**1243–1247, 1998.
269. Huntington Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72:**971–983, 1993.
270. McNeil, S., Novelletto, A., Srinidhi, J., Reduced penetrance of the Huntington's disease mutation. *Hum. Mol. Genet*. **6:**775–779, 1997.
271. Kremer, B., Almqvist, E., Theilmann, J., et al. Sex-dependent mechanisms for expansions and contractions of the CAG repeat on affected Huntington disease chromosomes. *Am. J. Hum. Genet*. **57:**343–350, 1995.
272. Rosenblatt, A., Brinkman, R., Liang, K., et al. Familial influence on age of onset among siblings with Huntington disease. *Am. J. Med. Genet*. **105:**399–403, 2001.
273. Bird, T. Outrageous fortune: the risk of suicide in genetic testing for Huntington disease. *Am. J. Hum. Genet*. **64:**1289–1292, 1999.
274. Meiser, B. and Dunn, S. Psychological impact of genetic testing for Huntington's disease: an update of the literature. *J. Neur. Neurosurg. Psychiatry*. **69:**574–578, 2000.
275. Craufurd, D., Kerzin-Starrar, L., Dodge, A., and Harris, R. Uptake of presymptomatic predictive testing for Huntington disease. *Lancet* **21:**603–605, 1989.
276. Meissen, G., Myers, R., Mastromauro, C., et al. Predicitive testing for Huntington's disease with use of a linked DNA marker. *N. Engl. J. Med*. **318:**535–542, 1988.
277. Hannig, V. L., Hopkins, J. R., Johnson, H. K., Phillips, J. A, and Reeders, S. T. Presymptomatic testing for adult onset polycystic kidney disease in at-risk kidney-transplant donors. *Am. J. Med. Genet*. **40:**425–428, 1991.
278. Babul, R., Adam, S., Kremer, B., et al. Attitudes toward direct predicitive testing for the Huntington disease gene: relevance for other adult-onset disorders. *J. Am. Med. Assoc*. **270:**2321–2325, 1993.
279. Farrer, L. Suicide and attempted suicide in Huntington disease: implications for preclinical testing of persons at risk. *Am. J. Med. Genet*. **24:**305–311, 1986.
280. Harper, P. *Huntington's Disease,* 2nd ed., W.B. Saunders, London, 1996, pp. 85–93.
281. International Huntington Association and the World Federation of Neurology Research Group on Huntington's Chorea. Guidelines for the molecular genetics predictive test in Huntington's disease. *J. Med. Genet*. **31:**555–559, 1994.
282. Huggins, M., Bloch, M., Wiggins, S. et al. Predictive testing for Huntington disease in Canada: adverse effects and unexpected results in those receiving a decreased risk. *Am. J. Med. Genet*. **42:**508–515, 1992.
283. Almqvist, E. W. A worldwide assessment of the frequency of suicide, suicide attempts, or psychiatric hospitalization after predictive testing for Huntington disease. *Am. J. Hum. Genet*. **63:**1293–1304, 1999.
284. Sermon, K., De Rijcke, M., Lissens, W., et al. Preimplantation genetic diagnosis with exclusion tesing. *Eur. J. Hum. Genet*. **10:**591–598, 2002.
285. Harper, P. and Sarfazi, M. Genetic prediction and family structure in Huntington's chorea. *Br. Med. J*. **290:**1929–1931, 1985.
286. Stern, G., Harton, G., and Sisson, M. Nondisclosing preimplantation genetic diagnosis for Huntington disease. *Am. J. Hum. Genet*. **67:**42a, 2000.
287. Berk, T., Cohen, Z., Bapat, B., and Gallinger, S. Negative genetic test result in familial adenomatous polyposis—clinical screening implications. *Dis. Colon Rectum* **42:**307–310, 1999.
288. Peshkin, B. N., Schwartz, M. D., Issacs, C., Hughes, C., Mail, D., and Lerman, C. Utilization of breast cancer screening in a clinically based sample of women after BRCA 1, 2 testing. *Cancer Epidemiol Biomarkers Prev*. **11:**1115–1118, 2002.
289. Botkin, J. R., Smith, K. R., Croyle, R. T., et al. Genetic testing for a BRCA1 mutation: Prophylactic surgery and screening behavior in women 2 years post testing. *Am. J. Med. Genet*. **118A:**201–209, 2003.
290. Michie, S., Collins, V., Halliday, J., and Marteau, T. Likelihood of attending bowel screening after a negative genetic test: the possible influence of health professionals. *Genet. Test*. **6:**307–311, 2002.
291. Bryant, L., Murray, J., Green, J., Hewison, J., Sehmi, I., and Ellis, A. Descriptive information about Down syndrome: a content analysis of serum screening leaflets. *Prenat. Diagn*. **21:**1057–1063, 2001.
292. Johnson, K. A, Trimbath, J. D., Peterson, G. M., and Giardiello, F. M. Impact of genetic counseling and testing on colorectal cancer screening behavior. *Genet. Test.* **6:**303–306, 2002.

44 Ethical, Social, and Legal Issues Related to Molecular Genetic Testing

JENNIFER L. HERBST AND JON F. MERZ

1. INTRODUCTION

Although a single chapter cannot address the full panoply of legal and ethical concerns associated with diagnostic genetic testing, this chapter examines three of the more pressing issues. We start by discussing informed consent of clinical patients and research subjects for genetic testing and research. We then move on to the confidentiality of genetic information and particularly concerns raised by the potential for discrimination or stigmatization of individual patients or their family members associated with disclosure of this information. We finish up with an overview of the implications for molecular pathologists posed by the patenting of genes.

Before launching into these discussions, it is important to recognize that genetic information differs from other medical information in a number of significant ways. Whereas most medical information is relevant only to the individual patient, genetic information (especially in the form of a predictive genetic test) can have direct, substantial implications for third parties, specifically family members and future generations *(1)*. In this sense, a genetic disease is akin to an infectious disease. The potential for conflict between the interests of the potentially affected third parties and the interests of his or her relatives is significant in health care and legal systems based heavily on patient autonomy *(2)*. However, unlike bacterial or viral diseases, genetic diseases are only transmitted through reproduction *(3)*. Consequently, any attempt to prevent the transmission of these diseases necessarily implicates the reproductive decisions of carriers. Reproductive decisions have been ethically and legally recognized as among the most intimate and sacrosanct and certainly the most deserving of privacy *(4,5)*. Not surprisingly, the technologies currently used to prevent the transmission of genetic disease from one generation to the next invoke some of the more debated and controversial technologies in bioethics, including in vitro fertilization, preimplantation diagnosis, and abortion.

Genetic data are unique in how they could affect identity and self-perceptions *(3)*. With the hype and publicity surrounding the sequencing of the human genome and the future of genomic medicine, the public arguably has been led to believe that we are little more than a product of our genes. For many, their genetic makeup is a key component of their identity. Knowledge of a mutation can greatly affect an individual's perception of his/her identity *(2)*. With current technology, the characteristics of a person revealed through genetic information are fixed—unchanging and unchangeable *(6)*. Surgery, drug therapies, and lifestyle changes can, in many cases, significantly change an individual's risk for development of other types of medical conditions; however, these modifications of diet, exercise patterns, jobs, location, or medical precautions might be perceived as being less effective against our predetermined genetic destinies. Information to be gained from a patient's DNA is unique to each individual and, thus, is inherently identifiable with current technology. Although anonymized tissue samples and DNA sequences are difficult to trace back to specific individuals, the possibility still exists, unlike with other kinds of medical information.

This chapter has been written with these specific properties of genetic information in mind. To address each of our three topic areas, we pose practical questions that might arise in daily activities of a molecular pathology laboratory.

2. INFORMED CONSENT

Before any clinical genetic testing or research can take place, a patient or subject must part with the tissue to be tested or studied. The term "informed consent" refers to the process during which a clinical patient or research subject agrees to surrender his/her tissue only after learning and understanding the reasonably foreseeable risks and benefits of doing so. An informed consent is a patient's or subject's acceptance of these foreseeable risks. The concept of an informed consent has played a large role in legal and bioethical thinking because of its potential for protecting and advancing several principles: respect for patients and subjects, protection of individuals' autonomy and self-determination, and the belief that it is wrong to force a person to act against his or her will. In the realm of

From: *Molecular Diagnostics: For the Clinical Laboratorian, Second Edition*
Edited by: W. B. Coleman and G. J. Tsongalis © Humana Press Inc., Totowa, NJ

research, informed consent is also a check on the power of investigators to perform unethical, morally suspect, or undesirable experiments.

For purposes of this discussion, we envision that molecular pathologists play both a clinical role in providing testing services for patients with whom they have no contact, as well as a research role in developing genetic tests and methods, validating tests with excess patient samples, and performing tests (using existing genes or searching for new genes) as part of formal research protocols. To make the distinction between clinical testing and research clear, when we refer to clinical testing, we are referring to the delivery of an individual's diagnostic genetic testing results to that individual, either through a physician or directly to the tested individual or an individual giving a tissue sample to a testing laboratory for an affirmative or negative confirmation of the presence of a specific mutation in his/her DNA. Research can take place with clinical collaborators who collect blood and other tissue samples from subjects for study or it can be performed on archived samples. Sometimes, research might also be performed on clinical samples at the same time or shortly after they are processed for clinical purposes. The primary beneficiary of the knowledge gained through clinical genetic testing is the patient. The primary beneficiary of knowledge gained through research is the scientific community. Regardless of whether testing is done for clinical or research purposes, each of these situations raise unique informed consent issues.

2.1. WHAT IS INFORMED CONSENT FOR CLINICAL TESTING? The duty of physicians to secure consent from patients arose out of the law of battery, which protects individuals from unwanted bodily contact or touching *(7)*. Consent, broadly speaking, is permission of the patient for his or her doctor to provide care. Informed consent, which evolved in a loose sense from this body of law, protects the interest of patients in having information provided so that they can contribute in a meaningful way to decisions affecting their medical care. Informed consent is the ethical and legal doctrine that requires health care providers to disclose information about what is proposed to be done, what the risks and the potential benefits are of the proposed course of treatment, and what the medically accepted alternatives are to their patients before treating them. Ethically, because patients do not have the specific knowledge to make decisions and the goal is to enable them to make decisions consistent with their own life experiences, goals, and expectations, health care providers should try to provide all information the individual patient would want and need to exercise this self-determination. Legally, such a subjective standard has not been adopted anywhere in the United States or in the United Kingdom, but it has been adopted in Canada.

Currently, most state laws tend to fall under one of two standards of disclosure—that of the "reasonably prudent physician" or that of the "reasonably prudent patient." A standard based on the reasonably prudent doctor requires disclosure of only that information that other doctors would typically convey to patients under the same circumstances. The reasonably prudent patient standard requires that doctors disclose all information that would be considered "material" by a reasonably prudent patient for making treatment decisions. This "reasonable disclosure of risks" is not limited to the physical risks to a patient, but has been extended to include risks of nontreatment or refusal of treatment *(8)*, risks specific to a particular physician or treatment center *(9)*, and physician conflicts of interest *(10)*.

This last concern, that of a physician's conflict of interest, was the subject of a lawsuit filed against the Regents of the University of California. In that case, the plaintiff, John Moore, underwent treatment for hairy-cell leukemia at the Medical Center of the University of California at Los Angeles. During his treatment, his physicians decided to use Mr. Moore's tissue samples (blood, blood serum, skin, sperm, bone marrow aspirate, and parts of his spleen) for research activities unrelated to Moore's medical care. The physicians did not inform Mr. Moore about this research. Eventually, one of the physicians established, patented, and licensed a cell line from Mr. Moore's T-lymphocytes. Moore sued the physicians for their failure to disclose the extent of their research and economic interests in Mr. Moore's cells before obtaining consent to the medical procedures by which the cells were extracted. The court found in favor of Mr. Moore, holding that the physician violated Mr. Moore's right to informed consent. Importantly, the Supreme Court of California also held that Moore had no property interest in the tissues collected by his doctors.

2.2. AM I RESPONSIBLE FOR GETTING INFORMED CONSENT FROM PATIENTS FOR CLINICAL TESTING? At present, molecular pathologists have no duty to get consent before performing a diagnostic test on a tissue sample submitted by a treating physician. Informed consent typically arises out of the direct interpersonal relationship between health care providers and patients or researchers and subjects. It is reasonable for laboratorians to assume that all submissions of samples are the result of an informed consent between the treating physician and patient. However, third parties, such as those involved in surgical pathology, diagnostic radiology, and diagnostic testing laboratories, often have a better understanding than a treating physician of the relevant information needed to make an informed decision about the specific test. This understanding makes these third parties better situated to provide necessary information to patients and subjects. In compiling its 2000 report on genetic testing, the Secretary's Advisory Committee on Genetic Testing for the National Institutes of Health came across a frequently mentioned need for health care providers to demonstrate competence in understanding genetic information and its implications. Indeed, one laboratory stated that "many of the questions we receive from client health care providers and patients relate to the translation and interpretation of genetic information in our medical reports." Comments also mentioned that inadequate public understanding and physician education cause much of the confusion and many of the risks associated with genetic testing (*11*; see also Secretary's Advisory Committee on Genetic Testing, 2000, Enhancing the Oversight of Genetic Tests: Recommendations of the SACGT, http://www4.od.nih.gov/oba/sacgt/reports/oversight_report.htm). The ultimate goal of any informed consent process should be to provide this education. Because clinical laboratorians possess the best information about the diagnostic genetic tests, testing laboratories have a unique gate-

keeper role that can be used to aid education of both clinicians and patients through informed consent procedures.

One possible means of passing on the necessary information to physicians and patients is to encourage and refer patients to genetic counselors. Another option for educating patients and physicians is to tap into the (hopefully) pre-existing infrastructure for informed consent by providing a consent form for clinicians to use when collecting blood for testing. An example consent form provided by one academic laboratory is included as an appendix. If a diagnostic genetic testing laboratory conditions testing on receipt of a signed consent with each tissue sample, patients are more likely to receive accurate information regarding the limitations of a particular test. Health care providers and patients will also be aware of the laboratory's tissue storage and confidentiality protocol before the tissues are submitted for testing.

2.3. IF I DECIDE TO USE AN INFORMED CONSENT FORM FOR CLINICAL GENETIC TESTING, WHAT INFORMATION SHOULD I PROVIDE IN THE FORM? Questions that should be addressed in a Clinical Informed Consent form include the following:

- How is my genetic material collected?
- How is my blood/skin/tissue tested?
- What will this test tell me?
- Do I have to take this test?
- What alternatives are there to taking this test?
- What are the risks of this test?
- How will I possibly benefit from this test?
- How will I learn the results of this test?
- How reliable are these results?
- How will this information about me be kept private?
- Who can have access to my results?
- To whom will my results be disclosed?
- What will happen to my sample after the test is completed?
- Will my sample be used for anything other than this diagnostic genetic test (e.g., research)?
- Who do I contact if I have questions or concerns about this test?

Patients should be informed about the length of time their information will be stored and the circumstances under which it would be expunged. They should also be provided with a clear statement about future secondary uses of the information, including quality assurance and quality improvement activities, as well as research *(6)*.

3. INFORMED CONSENT FOR RESEARCH

Although informed consent for clinical medical care, including diagnostic pathology, is regulated primarily through state law, there are US federal rules governing human subjects research that apply to research activities with human tissues. The regulations are referred to as the Common Rule because they have been adopted by all federal agencies that perform or sponsor human subjects research *(12)*. Broadly, these rules require informed consent of subjects and that all research be approved by an Institutional Review Board (IRB). The rules technically apply to federally funded research, but many large institutions apply the rules across the board to all research, and the standards of informed consent and prior IRB approval are truly the standard of care of human subjects research in the developed world. The Common Rule was promulgated with an eye toward reducing the burden on IRBs by minimizing the process of review for research deemed to present little or no risk to subjects. In addition, the Food and Drug Administration has its own set of rules applicable to clinical trials performed in support of applications for licenses for drugs and devices *(13)*. The protections offered subjects under these rules as applied by IRBs are highly similar, as IRBs rarely attend to the specific source of a legal or ethical requirement when making judgments about human subjects protections.

3.1. WHAT IS HUMAN SUBJECTS RESEARCH? The regulations define research as a systematic investigation designed to develop or contribute to generalizable knowledge, and they generally exclude activities related to quality assurance *(14)*. Activities related to clinical test development and validation are probably not research, but they raise concerns relating to test results and identifiable patients. Human subjects are defined as any living individual about whom an investigator conducting research obtains (1) data through intervention or interaction with the individual or (2) identifiable private information *(15)*. In this case, the information is identifiable if the identity of the subject is or could readily be ascertained by the investigator or associated with the information, and the information is considered to be private if it represents that which subjects have a reasonable expectation will not be made public. Thus, if an investigator gains access to human tissues or medical information and has the ability to identify any individual, then his or her research is human subjects research. The rules require all human subjects research to be approved by an IRB and, subject to waiver under certain circumstances, to have the informed consent of subjects.

3.2. WHAT NEEDS TO BE INCLUDED IN A RESEARCH INFORMED CONSENT? To comply with the Common Rule, informed consent for a research protocol requires disclosure of the information listed below *(16)*. True informed consent requires that a subject understand all this information before he/she agrees to participate in a research study.

- An explanation of the purposes of the research
- The expected duration of the subject's participation
- A description of the procedures to be followed
- Identification of any experimental procedures
- A description of the reasonably foreseen risks and benefits to the subject
- Disclosure of appropriate alternative procedures or courses of treatment, if any, that might be advantageous to the subject
- A statement describing the extent to which confidentiality of records identifying the subject will be maintained
- A statement that participation is voluntary
- An explanation of whom to contact for answers to pertinent questions about the research and research subjects' rights

Informed consent for research will differ from protocol to protocol. Procedures, language, and information previously developed and recommended for family-based research are not well suited to most population-based studies of low-penetrance gene variants. These recommendations generally

fail to distinguish between studies expected to reveal clinically relevant information about participants and studies expected to have meaningful public health implications but that involve few physical, psychological, or social risks for the individual subject *(1)*.

3.3. CAN I PERFORM RESEARCH ON TISSUE SAMPLES WITHOUT INDIVIDUAL CONSENT? There are two primary exceptions under which research using human tissue and medical records can proceed without the consent of subjects. First, use of human tissues in research can be exempt from the federal consent requirements if the tissues have already been obtained (for other purposes) prior to the time the research is proposed and the researcher (or preferably someone else) records the information in an anonymized form (i.e., with no identifying data or linking codes that would permit reidentification of an individual) *(17)*. However, the extent to which genetic information can truly be anonymized is currently unclear. Second, research (using even identifiable tissues) can receive expedited review (approval by only the IRB chair or a designee), and informed consent can be modified or waived completely if the IRB determines and documents that the research poses no more than minimal risk to subjects, that the waiver will not adversely affect subjects' rights or welfare, and the research could not practicably be carried out without the waiver *(18)*. IRB waiver of consent is made easier with stringent confidentiality policies because it reduces the risks to subjects' rights and welfare.

4. CONFIDENTIALITY AND DISCRIMINATION

Once patients and subjects have consented to genetic testing or a waiver of consent is granted by an IRB, the confidentiality of any information gleaned from this testing becomes an important concern. Information is considered confidential when it is meant to be kept secret. A confidence is placed in a professional who possesses personal information, such as a database of DNA sequences or genetic testing results, for the benefit of others. Medical information has long been treated as confidential because this confidentiality plays an important part in providing health care. Patients are less likely to divulge sensitive information to health professionals if they are not assured that their confidences will be respected *(6)*.

New regulations issued under the federal Health Insurance Portability and Accountability Act (HIPAA) were implemented in early 2003. HIPAA rules require that the IRB or a privacy board find and document that the following conditions are satisfied *(19)*:

- The use or disclosure of protected health information involves no more than minimal risk to the individuals.
- The alteration or waiver will not adversely affect the privacy rights and the welfare of the individuals.
- The research could not practicably be conducted without the alteration or waiver.
- The research could not practicably be conducted without access to and use of the protected health information.
- The privacy risks to individuals whose protected health information is to be used or disclosed are reasonable in relation to the anticipated benefits, if any, to the individuals, and the importance of the knowledge that might reasonably be expected to result form the research.
- There is an adequate plan to protect the identifiers from improper use and disclosure.
- There is an adequate plan to destroy the identifiers at the earliest opportunity consistent with conduct of the research, unless there is a health or research justification for retaining the identifiers, or such retention is otherwise required by law.
- There are adequate written assurances that the protected health information will not be reused or disclosed to any other person or entity, except as required by law, for authorized oversight of the research project, or for other research for which the use or disclosure of protected health information would be permitted by the (rule).

In addition to its utilitarian role in maintaining effective health care systems, confidentiality also serves the patients' (and subjects') interests in and rights to personal privacy. Westin defined the "core concept of privacy" as "the claim of an individual to determine what information about himself or herself should be known by others" *(20)*. Unwanted or unjustified disclosure of an individual's personal information can seriously compromise that individual's dignity.

Under the broad heading of personal privacy, a more specific interest in genetic privacy is evolving. The term "genetic privacy" is used to describe informational privacy, including the confidentiality, anonymity, or secrecy of the data that result from genetic testing and screening, sampling, and research *(20)*. The primary issues raised concern the genetic information contained in stored tissues *(21)*. Exploding genetic technology has expanded our ability to read and understand this information. Nonetheless, this information is not value neutral—it can be important to some people, totally unwanted by others, and harmful and injurious if disclosed or misused. Issues of access to and use of previously stored tissues have been raised, as have concerns about obtaining informed consent for storing clinical samples and the use of these samples for future research *(22–25)*.

Maintenance of individuals' genetic privacy is important within the context of diagnostic genetic testing because the greatest risks associated with genetic testing are not physical or medical risks to the patient or subject, but discrimination and other social and psychological harms *(1)*. Traditionally, wrongful discriminatory actions have been based on race, age, gender, nationality, mental or physical handicap, and ethnicity—things that, like genetic information, are unchanging and unchangeable. In the past, diagnoses of diseases such as diabetes and cancer have been stigmatizing. The public perceives an individual's genetic makeup as a type of blueprint of their essence *(20)*. The single-nucleotide polymorphisms associated with genetic diseases or an increased risk of disease are deviations from the norm—more specifically, something other than what it should be (to be an otherwise "healthy" person). A person with one of these markers might be seen as defective, especially when an effective treatment has not yet been developed *(20)*. As a result, individuals with disease markers might be stigmatized and subjected to discrimination *(26–28)*. Specific examples of this discrimination might surface in employment

decisions such as hiring, firing, or promotion *(27)*. Long-term goals and plans such as those of adoption, advanced education, or home ownership might be frustrated if a risk of disability or illness is disclosed *(29,30)*. One important factor in assessing the probability and magnitude of these harms is the potential clinical relevance of the results *(1)*. When assessing possible harms, the National Bioethics Advisory Commission suggested considering these questions *(31)*:

1. How easily identifiable is the source of the tissue sample?
2. What is the likelihood that the source will be traced?
3. If the source is traced, what is the likelihood that persons other than the investigators will obtain information about the source?
4. If noninvestigators obtain information regarding the source, what is the likelihood that harms will result, including adverse consequences arising from the reporting of uncertain or ambiguous clinical results?

The current legal framework for privacy protection is fragmented and inconsistent *(32)*. Individual genetic information has very little protection under federal law. Although the United States' Constitution (as interpreted by the Supreme Court) does grant American citizens a fundamental right to privacy, this right only protects citizens from government action, not from actions by private parties. Under HIPAA, group health plans cannot deny eligibility based on presymptomatic genetic information, a term that the statute fails to define *(33)*. HIPAA prohibits the use of genetic information to deny or limit insurance coverage for members of a group health plan. Additionally, a presymptomatic diagnosis cannot be considered a pre-existing condition. However, these protections in HIPAA are seriously limited. First, HIPAA only applies to group health plans—not the individual health insurance market. Individuals interested in insurance coverage outside the group market can still be denied coverage or charged higher premiums based on genetic test results. Second, although HIPAA prevents group plan insurers from treating similarly situated individuals differently (i.e., charging different premiums to two individuals with similar health profiles), there is nothing (other than competitive market pressures) preventing an insurer from charging all members of a group a higher premium based on the genetic information of one individual in the group. Third, HIPAA does nothing to limit access or release of genetic information to insurers. There is no federal law that prohibits individual insurance companies from conditioning coverage on a genetic screening or access to test results from previous genetic tests that are contained in medical records or family histories *(20)*. Many states have enacted legislation in an attempt to fill in some of the holes left by HIPAA and federal constitutional law, trying to prevent discrimination based on genetic information *(34)*. As is usual with state-by-state legislative efforts, these have failed to provide a comprehensive solution to discrimination in health insurance *(20,35,36)*.

There is a strong belief in the legal literature discussing genetic privacy that authorized users of health information pose a greater threat to individual privacy than unauthorized users *(6)*. Authorized use of health information extends far beyond those who directly deliver or manage patient health care or financial reimbursement. Patient records (and the genetic information found therein) can be accessed lawfully by parties interested in education (e.g., conferences, teaching hospitals, and continuing education), regulation (e.g., litigation and accreditation), commercialization (e.g., development of research and marketing strategies), social services and child protection (e.g., tracking and intervening in spousal or child abuse), and public health (e.g., disease reporting, partner notification, and disease or postmarketing drug and device surveillance) *(6)*. Additionally, the state has an interest in the well-being of its citizens *(37)*. The government can trump an individual's right to privacy if the program is narrowly tailored to serve this legitimate state interest *(38–40)*. One example of such a policy is a newborn screening statute to detect and treat a specific genetic disease such as phenylketonuria (PKU). This particular test has been especially successful because the screening process is reliable and accurate, involves a readily identifiable population, and detects a condition that has an effective treatment available to those affected *(41)*.

In light of these legal shortcomings and potential harmful effects of inappropriate disclosure, genetic testing laboratories should have a very high level of protection of genetic information. Enhanced privacy protections can include some or all of the following. First, research projects should provide more protection than clinical testing because insurance companies are not paying for the testing and test results can be kept out of official medical records *(2)*. Additionally, clinical patients might be informed of potential discrimination as part of the informed consent process and presented with the option of paying for the test out-of-pocket rather than through their insurance companies. Providers might also use pseudonyms or anonymizing codes to protect patients' identities. Of course, protecting patient privacy in these ways risks undermining care, and clinicians should assure themselves that they can adequately monitor and treat their patients under these testing scenarios. Some of the practices currently used to protect patient or subject confidentiality include (1) keeping linking files physically and electronically separate from data files, (2) destructing identifying information or inking files as soon as no longer needed to accomplish the research purpose for which data are collected or collaged, (3) collecting of only anonymous data, meaning that no one associated with the research could directly or by use of a code identify an individual from the research data, and (4) involving clinic staff in data collection activities such that only individuals having clinical responsibilities would have access to patient identities, and securing written confidentiality assurances from non-health-care profession research staff *(42)*.

The following groups of questions are some that clinical laboratorians need to consider when developing research protocols as well as sample use and storage policies to identify possible risks:

1. In any protocol involving collection of human tissues, will there be any excess tissues left over that will not be destroyed or discarded? If so, who will keep them? Where will they be stored? For how long? Will samples be identified or linked to other data about individuals that would permit identification of the subject?
2. What will happen to information potentially developed in the current research? If subjects are identifiable, have they consented to receiving or not receiving any information

developed? If they have consented to receive information, is that information likely to be such that prior counseling should be provided about the information? Who will pay for the counseling? If subjects have consented to not receive information, are potentially identifying links being maintained? If so, for how long, and for what purposes?

3. Have subjects consented expressly to the storage of excess tissues and use in research?
4. For planned use of stored tissues, are identifying links necessary to perform the research? If not, deletion of these links should be required. If so, then plans for contacting persons whose tissues are sought to be used and securing consent must be approved. Who will make contact? Have potential subjects given consent to such contact?

5. GENE PATENTING

Numerous laboratorians in the United States report receiving letters from patent holders, informing them that particular tests have been patented and that the laboratory, if performing one of the tests, must either pay a royalty for the right to continue doing so or must cease performing the test *(43)*. In the latter case, the laboratory is told where patient samples can be sent to have the patented tests performed and how much it will cost. Being compelled to stop providing testing services has serious implications for the ability of molecular pathologists to maintain currency in their field, to treat their patients with comprehensive medical services, to train residents and fellows, to perform research, and to run their labs in an efficient manner. In the final analysis, these patents restrict physicians' ability to practice medicine *(44,45)*. Despite these concerns, laboratorians must deal with patents and patent owners in a responsible manner.

5.1. WHAT ARE PATENTS? Utility patents are granted by the federal government to inventors for new, nonobvious, useful inventions and discoveries. A patent grants to its owner the right to exclude others from making, using, or selling a patented machine or composition of matter, or using a patented method. United States patents, like those throughout the world, currently are enforceable for 20 yr from the date of filing a patent application. A patent owner generally is not under any legal compunction to work or utilize a patented invention.

A patent grants only a negative right—the right to enjoin others from using the invention without permission—and a patent owner can sue in federal court any infringer to block them from making, using, or selling the invention. Monetary damages equivalent to a reasonable royalty plus costs can be recovered, and treble damages might be awarded by a court for willful infringement. Patent litigation is extremely expensive and, fortunately, relatively rare *(46)*.

Patents relating to molecular diagnostics cover primers and other chemical compounds and mixtures useful in diagnostics, machines such as sequencers, and methods of using these. Further, patents have been awarded for the discovery and characterization of specific genes, as well as for useful DNA fragments. These patents typically cover isolated and purified compositions of matter (including cDNA, protein, antibodies, and cell lines), potential products made using the sequence (such as a recombinant protein or a gene transfer vector containing the sequence), as well as diagnosis of any disclosed disease-related mutations or polymorphisms in the gene. Cook-Deegan and his colleagues estimate that some 28,000 US gene patents have been issued to date (Cook-Deegan, personal communication).

Patents that might be familiar to laboratorians include those covering polymerase chain reaction (PCR) and related technologies. Typically, when a lab purchases PCR machines and reagents, the prices include a premium paid for the exclusivity granted by the patents. For PCR, an additional royalty of about 9% has been reportedly been paid for all testing done by licensed laboratories *(47)*. A more recent example are patents that claim the use of intronic sequences for generating haplotypes and identifying allelic variation (US Patents 5,851,762, 5,612,179, 5,192,659, and 5,096,557 are being enforced by an Australian firm, Genetic Technologies, Ltd., which acquired the patents from GeneType, AG; see http://www.gtg.com.au/). Untold genes have also been patented, and many patent applications are still pending (during which time they are often kept secret). Well-known examples include those covering genes implicated in breast and ovarian cancers (BRCA1 and BRCA2), colon cancers (HNPCC, FAP), cystic fibrosis (CFTR), hemochromatosis (HFE), and a growing number of neurological diseases including late-onset Alzheimer's disease (Apo-E), Canavan disease, Charcot–Marie–Tooth disease (CMT-1A and CMT-X), spinal muscular atrophy (SMN1), spinocerebellar ataxia (SCA1-SCA12), and others.

5.2. DO I HAVE TO STOP PERFORMING ANY PATENTED TEST? For the majority of gene patents issued to date, the patent owners are not aggressively enforcing their rights against clinical molecular diagnostics laboratories. Nonetheless, most clinical laboratories have received letters from different patent owners informing them about particular gene patents *(43)*. In some cases, these patent owners have been willing to grant a license to laboratories performing a home-brew test. Per-test royalties of which we have become aware include $2 for the ΔF508 mutation of CFTR (University of Michigan), $5 for Gaucher's Disease (Scripps Institute), $12.50 for Canavan disease (Miami Children's Hospital), and reportedly more than $20 for HFE (Bio-Rad). In some cases, an up-front license fee has been demanded as well *(48)*. In a small number of instances, patent owners have not been willing to provide licenses to laboratories to perform the tests, instead deciding to monopolize the testing service and requiring labs to send samples for testing to the owner. Thus, tests for breast and ovarian cancer genes (Myriad Genetics) and a list of neurological disorders (Athena Diagnostics) are generally available from only these commercial laboratories.

It is unreasonable to expect that laboratorians will follow all of the voluminous patenting activity in molecular biology and be responsible for knowing the patent landscape in which they work. Further, patent owners must be diligent in enforcement of their patent rights, and failure to try to determine who is infringing and taking steps to notify potential infringers in order to stop or license them might result in their inability to later enforce the patent. A 6-yr delay in doing something about known infringement will bar later enforcement, although this time might be shorter if a patent holder's delay results in material prejudice to an infringer *(49)*.

If laboratories receive notice from a patent owner of patents they are or might be infringing, they could try to negotiate a license, they could stop using methods that infringe, or they could choose to ignore the letter and take the risk that they will be sued. To date, only one academic laboratory in the United States is known to have been sued by a commercial gene patent owner, and that institution quickly stopped performing the test. In any event, legal counsel should be sought if any course other than dropping the test is contemplated.

5.3. WHAT IF I AM PERFORMING TESTING FOR RESEARCH; IS THIS EXEMPT FROM PATENT INFRINGEMENT? Although many people believe that research does not infringe patents, this is not strictly true. There is an extremely limited research exemption in United States patent law. As recently stated by the Court of Appeals for the Federal Circuit in a suit against Duke University,

> regardless of whether a particular institution or entity is engaged in an endeavor for commercial gain, so long as the act is in furtherance of the alleged infringer's legitimate business and is not solely for amusement, to satisfy idle curiosity, or for strictly philosophical inquiry, the act does not qualify for the very narrow and strictly limited experimental use defense. *(50)*

In the Madey case, Duke University's use of infringing laboratory equipment for research and educational purposes did not excuse the university from patent liability *(50)*.

In another recent case, a group of collaborative researchers, the Cancer Genetics Network (CGN), funded by the US National Cancer Institute (NCI), was threatened because of their infringement of patents held by Myriad Genetics on BRCA1 and BRCA2 (http://dccps.nci.nih.gov/cgn/default.html). The fact that genetic testing of subjects—involving the sequencing of the two patented genes—was performed within a research protocol offered no protection from liability to these academic centers. Thus, the researchers were essentially compelled by Myriad to use their testing services. To the company's credit, the NCI was able to negotiate a reduced price and other favorable terms for testing services (Memorandum of Understanding between Myriad Genetic Laboratories, Inc. and the National Cancer Institute, January 2000).

Although performing research in itself does not offer certainty of nonliability, molecular pathologists nonetheless have some realistic types of research they can do without excessive concern about infringement. For example, some would argue that clinical studies aimed at finding new genes, new mutations, new methods of testing, and better characterization of patented genes should be exempt. In this view, research aimed at better understanding of the claimed invention—such as how it works, whether it works as taught by the patent, how to improve upon it, and how to work around it—is desirable and should be promoted by the law. Although the scope of the exemption is uncertain, as a practical matter, research uses of patented genes and diagnostics are unlikely to receive much attention from a patent owner unless the studies are very large and the laboratories' practice of the patent is viewed as materially impinging upon the patent owner's commercial rights.

5.4. WHAT IF I DISCOVER A NEW GENE OR NEW MUTATIONS IN GENES; SHOULD I SEEK PATENT PROTECTION? In a world dominated by patents, owning patents assures one the ability to practice, provides a coin for negotiation with other patent holders, and gives the power to use an invention in a way the individual sees fit. The downside is that patents are relatively expensive to secure, costing upward of $10,000 for a US molecular biology patent, and perhaps 10 times that for a full range of international patents. The downside of not patenting is that one could, in the worst case, be later blocked from using the invention by patents on closely related discoveries and inventions. For example, if a researcher discovers a gene with several polymorphisms related to occurrence of a disease, others might later patent new polymorphisms that they discover in the same gene that have diagnostic utility. If the first discovery is not patented, the subsequent patentees could block the first researcher from diagnostic uses of the gene.

As an ethical matter, patenting by physicians has long been looked upon with derision, and for good reason *(51)*. Patents are particularly of concern for medical processes, covering methods such as those used by molecular pathologists. Patents restrict access, increase costs of services, and provide rewards for the practice of medicine by others. These concerns led to a 1996 federal law, named after its sponsors Ganske and Frist in the United States House of Representatives and Senate, respectively, that holds physicians and health care institutions immune from liability for infringement of pure process patents. Unfortunately, excepted from these protections are biotechnology patents and processes performed by CLIA-approved laboratories *(52)*. If patents are secured, they should be used in a responsible manner. This implies not restricting the use of the patented methods and not charging exorbitant licensing fees and royalties; patents covering diagnostic methods should be licensed nonexclusively to any laboratory that wishes to perform testing at a royalty, if any, that is no more than a small percentage of the marginal costs of performing the patented assay *(53)*. Technological advances are driving the cost of testing downward, and exclusive licenses and fixed royalties will delay the adoption of less expensive and faster technologies *(54)*.

6. CONCLUSION

There are many ethical issues inherent in the practice of clinical molecular pathology. We have tried to provide practical insights and guidance on some of the more common problems. Of course, the nature of the practice of medicine—particularly in a rapidly evolving field such as clinical genetics—is that new ethical problems and conundrums are sure to arise. Institutional ethics committees can be relied on for help in resolving tough issues as they arise.

ACKNOWLEDGMENT

The authors thank Dr. Debra G. B. Leonard for providing an example of an informed consent form for inclusion in this book chapter.

REFERENCES

1. Beskow, L. M., Burke, W., Merz, J. F., et al. Informed consent for population-based research involving genetics. *JAMA* **286:**2315–2321, 2001.
2. Brandt, J. Ethical considerations in genetic testing: an empirical study of presymptomatic diagnosis of Huntington's disease, in *Medicine and Moral Reasoning*, Fulford, K. W. M., Gillett, G. R., and Soskice, J. M., eds., Cambridge University Press, New York, 1994.
3. Suter, S. M. Whose genes are these anyway? Familial conflicts over access to genetic information. *Mich. Law Rev.* **91:**1854–1908, 1993.
4. Griswold v. Connecticut, 381 U.S. 479 (1965).
5. Roe v. Wade, 410 U.S. 113 (1973).
6. Gostin, L. O. Health information privacy. *Cornell Law Rev.* **80:** 480–528, 1995.
7. Furrow, B. R., Greaney, T. L., Johnson, S. H., Jost, T. S., Schwartz, R. L. *Health Law: Cases, Materials and Problems*, West Group, St. Paul, MN, 1999.
8. Truman v. Thomas, 611 P.2d 902 (Cal. 1980).
9. Johnson v. Kokemoor, 545 N.W.2d 495 (Wis. 1996).
10. Moore v. Regents of the University of California, 793 P.2d 479 (Cal. 1990).
11. Cho, M. K., Sankar, P., Wolpe, P.R., and Godmilow, L. Commercialization of BRCA1/2 testing: practitioner awareness and use of a new genetic test. *Am. J. Med. Genet.* **83:**157–163, 1999.
12. 45 CFR Part 46 (2002).
13. 21 CFR § 50.25 (2002).
14. 45 CFR § 46.102(d) (2002).
15. 45 CFR § 46.102(f) (2002).
16. 45 CFR § 46.116 (2002).
17. 45 CFR §46.101(b)(4) (2002).
18. 45 CFR § 46.116(d) (2002).
19. 45 CFR § 164.512(i)(2)(ii) (2002).
20. Turkington, R. C. and Allen, A. L. *Privacy Law: Cases and Materials*. West Group, St. Paul, MN, 1999.
21. Merz, J. F. On the intersection of privacy, consent, commerce and genetics research, in *DNA Sampling: Ethical, Legal and Social Issues*, Knoppers, B.M., ed., 2004.
22. Annas, G. J. Privacy rules for DNA databanks: protecting coded 'future diaries.' *JAMA* **270:**2346–2350, 1993.
23. Annas, G. J., Glantz, L. H., and Roche, P. A. *The Genetic Privacy Act and Commentary*. Boston University School of Public Health, Boston, MA, 1995.
24. Clayton, E. W., Steinberg, K. K., Khoury, M. J., et al. Informed consent for genetic research on stored tissue samples. *JAMA* **274:**1786–1792, 1995.
25. Deschenes, M., Cardinal, G., Knoppers, B. M., and Glass, K. C. Human genetic research, DNA banking and consent: a question of 'form'? *Clin. Genet.* **59:**221–239, 2001.
26. American Society of Human Genetics Board of Directors, American College of Medical Genetics Board of Directors. ASHG/ACMG report: points to consider: ethical, legal and psychosocial implications of genetic testing in children and adolescents. *Am. J. Hum. Genet.* **57:**1233–1241, 1995.
27. Billings, P. R., Kohn, M. A., deCuevas, M., Beckwith, J., Alper, J. S., and Natowicz, M. R. Discrimination as a consequence of genetic testing. *Am. J. Hum. Genet.* **50:**476–482, 1992.
28. Ostner, H., Allen, W., Crandall, L. A., et al. Insurance and genetic testing: where are we now? *Am. J. Hum. Genet.* **52:**565–577, 1993.
29. Wertz, D. C., Fanos, J. H., and Reilly, P. R. Genetic testing for children and adolescents: who decides? *JAMA* **272:**875–881, 1994.
30. Alper, J. S., and Natowicz, M. R. Genetic discrimination and the public entities and public accommodations: titles of the Americans with Disabilities Act. *Am. J. Hum. Genet.* **53:**26–32, 1993.
31. National Bioethics Advisory Commission. *Research Involving Human Biological Materials: Ethical Issues and Policy Guidance*. Volume 1. National Bioethics Advisory Commission, Rockville, MD, 1999.
32. Gostin, L. O. Genetic privacy. *J. Law Med. Ethics* **23:**320–330, 1995.
33. Annas, G. J. The limits of state laws to protect genetic information. *N. Engl. J. Med.* **345:**385–388, 2001.
34. Mulholland, W. F., and Jaeger, A. Genetic privacy and discrimination: a survey of state legislation. *Jurimentrics* **39:**317–326, 1999.
35. Hudson, K. L., Rothenberg, K. H., Andrews, L. B., Kahn. M. J., and Collins, F. S. Genetic discrimination and health insurance: an urgent need for reform. *Science* **270:**391–393, 1995.
36. Rothenberg, K. H. Genetic information and health insurance: state legislative approaches. *J. Law Med. Ethics* **23:**312–319, 1995.
37. Merz, J. F., McGee, G. E., and Sankar, P. "Iceland, Inc."? On the ethics of commercial population genomics. *Soc. Sci. Med.* **58:** 1201–1209, 2004.
38. Jacobsen v. Massachusetts, 197 U.S. 11 (1905).
39. Railroad Co. v. Husen, 95 U.S. 465 (1878).
40. Planned Parenthood of Southeastern Pennsylvania v. Casey, 505 US 833 (1992).
41. Burke, W., Atkins, D., Gwinn, M., et al. information needs of clinicians, policy makers, and the public. *Am. J. Epidemiol.* **156:**311–318, 2002.
42. Merz, J. F. Introduction: a survey of international ethics practices in pharmacoepidemiology and drug safety. *Pharmacoepidemiol. Drug Safety* **10:**579–581, 2001.
43. Cho, M. K., Illangasekare, S., Weaver, M. A., Leonard, D. G. B., and Merz, J. F. Effects of patents and licenses on the provision of clinical genetic testing services. *J. Mol. Diagn.* **5:**3–8, 2003.
44. Merz, J. F. Disease gene patents: overcoming unethical constraints on clinical laboratory medicine. *Clin. Chem.* **45:**324–330, 1999.
45. Leonard, D. G. B. Medical practice and gene patents: a personal perspective. *Acad. Med.* **77:**1388–1391, 2002.
46. Merz, J. F., and Pace, N. Trends in patent litigation: the apparent influence of strengthened patents attributable to the Court of Appeals for the Federal Circuit. *J. Patent Trademark Office Soc.* **76:**579–590, 1994.
47. Cho, M. Ethical and legal issues in the 21st century, in *Preparing for the Millenium: Laboratory Medicine in the 21st Century*. AACC, Orlando, FL, 1998.
48. Merz, J. F., Kriss, A. G., Leonard, D. G. B., and Cho, M. K. Diagnostic testing fails the test: the pitfalls of patenting are illustrated by the case of haemochromatosis. *Nature* **415:**577–579, 2002.
49. Ecolab, Inc. v. Envirochem, Inc., 264 F.3d 1358 (Fed. Cir. 2001).
50. Madey v. Duke University, 307 F.3d 1351 (Fed. Cir. 2002) (italics added).
51. American Medical Association. *Code of Medical Ethics: Current Opinions with Annotations*. American Medical Association, Chicago, IL ,1997.
52. US Code, title 35, § 287(c) (2003).
53. Schissel, A., Merz, J. F., and Cho, M. K. Survey confirms fears about licensing of genetic tests. *Nature* **402:**118, 1999.
54. Merz, J. F. Patents limit medical potential of sequencing. *Nature* **419:**878, 2002.

APPENDIX
AN EXAMPLE OF AN INFORMED CONSENT FORM FROM AN ACADEMIC LABORATORY

University Health System
Molecular Pathology Laboratory
Informed Consent for Genetic Testing

Please read and sign this form so that we can be sure you understand this genetic testing (also called "DNA testing") and the risks associated with it. Please ask questions about anything on this form you do not understand.

Purpose

I understand that a sample of blood will be drawn from me and/or members of my family for the purpose of attempting to determine if I and/or family members are carriers of the disease gene, or are affected with, or at increased risk to some day be affected with this genetic disease.

Background

DNA is a chemical that encodes hereditary information. Genes are specific pieces or subunits of DNA that have function in the body. Genes come in pairs, one from our mother and the other from our father. A DNA test can directly detect an abnormality, called a mutation. Mutations are most often found in the gene and result in abnormal gene function, which is associated with disease. Depending on the genetic condition a mutation in one gene may be associated with disease (autosomal dominant) or in other conditions (autosomal recessive) both genes need to have mutations to be associated with disease. The interpretation of my results will be explained to me when they are available.

Genetic Testing

This genetic test is specific for _____ (specific disease). There is a chance that I will have this genetic condition but the DNA test results will be negative. This is called a false negative. There is a ___% chance that this test will detect this condition if I have it. Due to laboratory techniques and capabilities some mutations that might be associated with this disease are not possible to test for at this time.

Some individuals who have an altered gene may never develop clinical symptoms associated with the condition. The chance for a gene mutation to cause symptoms has been discussed with me.

In some cases, an indirect DNA test called linkage analysis may have to be used. Linkage analysis involves blood samples from other specified family members. If linkage analysis is being used, naturally occurring rearrangements in the DNA, called recombination, may produce an uncertainty in the results. Linkage analysis is not 100% accurate because of the chance for recombination. Results from linkage analysis are reported as a probability.

In some families, the DNA markers used in testing may not be informative. If this is the case, this DNA test cannot provide results for that family or for some members of that family.

An error in the diagnosis may occur if the true biological relationships of the family members involved are not as I have stated. Also, testing may detect nonpaternity. Nonpaternity means that the father of an individual is not the person stated to be the father. DNA analysis is specific only with respect to the disease. It does not provide me with any information about the current status of my health. It in no way guarantees my health or the health of my unborn child or children.

DNA tests are relatively new and are being improved and expanded continuously. This testing is often complex and utilizes specialized materials so that there is always a small possibility that the test will not work properly or that an error will occur. There is an error rate, although low, even in the best laboratories. In the rare event of a laboratory error, I may be asked to have my blood redrawn. My signature below acknowledges my consent to have my blood redrawn in the event of laboratory error.

In some cases it may be possible for the laboratory to reanalyze leftover DNA samples in the future using new and improved testing methods. However, I understand that this is not a DNA banking facility and my DNA sample may not be available for future clinical studies.

Use of Specimens

I understand that any samples obtained for the purpose of this testing become the exclusive property of the Health System and that I relinquish all right, title and interest to such samples. I understand that my blood sample will only be used for DNA testing as authorized by my consent and that my DNA sample will not be made available for future clinical studies or research purposes without my consent.

Procedure

The procedure for drawing blood involves placing a needle in a vein in the arm to draw blood. This procedure only takes a few minutes. Occasionally there are minor complications, and I may experience bruising, swelling, black and blue marks, fainting and/or infection at the site. I have disclosed and discussed with the physician any pertinent information or concerns about a personal history of bleeding problems or any other medical conditions before having my blood drawn.

Alternatives

There may be other tests that my physician could use to help diagnose my condition. If I am reluctant to have the DNA test, I will discuss this with my physician.

Psychological Consideration

Sometimes DNA testing can have an adverse psychological effect on an individual. Some individuals may experience depression, increased anxiety, mood changes and other forms of emotional distress both before and after receiving results. However, some individuals report feeling a sense of relief, closure, and/or decreased anxiety from proceeding with DNA testing and receiving results.

(Continued)

APPENDIX *(Continued)*

Confidentiality
My results are confidential. Because DNA testing is complex and results occasionally difficult to interpret, my results will be reported to me only through a physician or genetic counselor who I designate. The medical information produced as a result of this testing will become part of my medical record and will only be released to other medical professionals or other parties with my written consent. However, my insurance carrier may receive information about my DNA testing if it is paying for my testing or treatment or if I have authorized the release of my medical records for other reasons. My signature below acknowledges my consent to the disclosure of my genetic testing results by the Health System to the insurance carrier(s) involved or to the individuals or entities listed below or as required by law. This disclosure may include test results, authorization or consent forms, chain of custody forms, or other documentation that may include personal health information.

Billing
The billing process for this DNA testing has been explained to me. I will be responsible for covering the costs associated with testing.

Discrimination Risks
Some individuals who have elected DNA testing and been found to have the gene leading to the disease have experienced discrimination. I understand that discrimination (insurance, employment, and social) is a risk of genetic testing, and that my health and life insurance rates, my ability to obtain health or life insurance and my employability could be affected.

Voluntary Consent
Participation in genetic testing is completely voluntary. When results are available I will be given the opportunity to decline hearing them. If I choose not to receive my results a record of the results will be kept in my chart for reference should I change my mind.

I will receive a copy of this consent form.

Patient Consent to Genetic Testing
I, _____ (print name), hereby agree to participate in testing for _____ (name of disease) using a DNA-based test. This testing will be performed by the Health System.
Patient Signature: _____
Date: _____

Authorization for Further Disclosure of Genetic Testing Results
I request that the Health System send copies of my genetic testing results to the following individuals and/or entities at the following addresses. I understand that the test results contain personal health information and that once my results are disclosed consistent with this authorization, federal or state privacy laws may no longer protect the health information contained in the results. I understand that I have the right to revoke this authorization in the future in a separate, signed writing indicating my intent to revoke. This authorization is valid as of the date of my signature below and shall remain valid for a period of twelve months beyond that date.
Send results to: (Provide full name and address of each individual to whom results should be sent.)

_____
_____
_____

Patient Signature: _____

Date: _____

Physician's/Counselor's Statement
I have explained DNA testing to_____. I have addressed the procedures involved the possible risks and benefits and the limitations outlined above, and I have answered this person's questions.
Physician/Counselor's
Signature: _____
Date: _____

INDEX

D

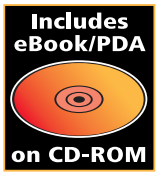
G

H

Q

R

S



Molecular Diagnostics

For the Clinical Laboratorian

SECOND EDITION

Edited by

William B. Coleman

Department of Pathology and Laboratory Medicine, University of North Carolina School of Medicine, Chapel Hill, NC

Gregory Tsongalis

Department of Pathology, Dartmouth Medical School,Dartmouth-Hitchcock Medical Center, Lebanon, NH

Foreword by

Lawrence M. Silverman

Department of Pathology and Laboratory Medicine, University of Virginia, Charlottesville, VA

From Reviews of the First Edition…

"…highly recommended…for those interested in clinical molecular diagnosis…will satisfy those who are more expert in certain areas of molecular diagnosis, but who have been unable to keep up with the wide range of recent advances in molecular diagnosis." ***—American Journal of Clinical Pathology***

"…gives a good theoretical background for those with little experience in this field. It is readable, comprehensive, and…up-to-date." ***—Journal of Clinical Pathology***

"…well written, comprehensive…should bring even the reader with limited background knowledge up to speed on what is a rapidly progressing field." ***—Trends in Molecular Medicine***

"…a timely and highly relevant book that will serve as a valuable resource to those practicing molecular diagnostics from the bench, as well as to basic and clinical scientists…" ***—Endocrine News***

"…a very good introduction to the basic elements of molecular pathology…excellent discussion of methods…" ***—Clinical Chemistry***

Taking advantage of the many major advances that have occurred since their groundbreaking first edition was published, William B. Coleman and Gregory J. Tsongalis have updated and expanded their highly praised tutorial guide to molecular diagnostic techniques to include not only improved traditional methods, but also totally new molecular technologies, some not yet in routine use. The authors offer cutting-edge molecular diagnostics for genetic disease, human cancers, infectious diseases, and identity testing, as well as new insights into the question of quality assurance in the molecular diagnostics laboratory. Additional chapters address other technologies found in the clinical laboratory that are complementary to molecular diagnostic methodologies, and also discuss genetic counseling and the ethical and social issues involved with nucleic acid testing. A value-added compact disk containing a companion eBook version of *Molecular Diagnostics: For the Clinical Laboratorian,Second Edition,* is included for downloading and use in the reader's PC or PDA.

Authoritative and state-of-the-art, *Molecular Diagnostics: For the Clinical Laboratorian,Second Edition* is the essential textbook of choice for anyone working in molecular diagnostics and who wants to remain current with this rapidly changing field.

Features

- **Cutting-edge molecular diagnostics for genetic disease, human cancers, and infectious diseases**
- **New applications of technology to specific problems in molecular diagnostics**
- **Discussion of quality assurance issues in the molecular diagnostics laboratory**
- **Advanced molecular technologies and their potential use in molecular diagnostics**
- **Inclusion of common technologies that can complement molecular diagnostics**
- **Introductory survey of basic molecular biology techniques and concepts**
- **Review of genetic counseling and the ethical issues involved in nucleic acid testing**

ISBN 1-58829-356-4

MOLECULAR DIAGNOSTICS: FOR THE CLINICAL LABORATORIAN, SECOND EDITION
ISBN: 1-58829-356-4 E-ISBN: 1-59259-928-1

humanapress.com