Molecular Diagnostics

Third Edition

Editor-in-Chief

George P. Patrinos

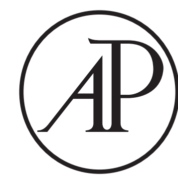
Associate Editors

Wilhelm J. Ansorge

Phillip B. Danielson

AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD • PARIS
SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier
125 London Wall, London EC2Y 5AS, United Kingdom
525 B Street, Suite 1800, San Diego, CA 92101-4495, United States
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom

Copyright © 2017, 2010, 2005 Elsevier Ltd. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Medical Disclaimer
Medicine is an ever-changing field. Standard safety precautions must be followed, but as new research and clinical experience broaden our knowledge, changes in treatment and drug therapy may become necessary or appropriate. Readers are advised to check the most current product information provided by the manufacturer of each drug to be administered to verify the recommended dose, the method and duration of administrations, and contraindications. It is the responsibility of the treating physician, relying on experience and knowledge of the patient, to determine dosages and the best treatment for each individual patient. Neither the publisher nor the authors assume any liability for any injury and/or damage to persons or property arising from this publication.

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-0-12-802971-8

For information on all Academic Press publications
visit our website at https://www.elsevier.com/

Publisher: Mica Haley
Acquisition Editor: Tari Broderick
Editorial Project Manager: Tracy Tufaga
Production Project Manager: Kirsty Halterman and Karen East
Designer: Greg Harris

Typeset by TNQ Books and Journals

List of Contributors

B.R. Ali United Arab Emirates University, Al-Ain, United Arab Emirates

W.J. Ansorge Ecole Polytechnique Federal Lausanne, EPFL, Lausanne, Switzerland

D.S. Atherton University of Alabama at Birmingham, Birmingham, AL, United States

M.P.G. Barnett AgResearch Limited, Palmerston North, New Zealand; Nutrigenomics New Zealand, Auckland, New Zealand

W.C. Bell University of Alabama at Birmingham, Birmingham, AL, United States

S.R. Brand Dana-Farber Cancer Institute, Boston, MA, United States

G. Cane Uppsala University, Uppsala, Sweden

W. Chantratita Mahidol University, Bangkok, Thailand

T. Conze Uppsala University, Uppsala, Sweden

D. Corach Universidad de Buenos Aires, Buenos Aires, Argentina

F. Coun KU Leuven, Leuven, Belgium

P.B. Danielson University of Denver, Denver, CO, United States; Center for Forensic Science Research and Education, Willow Grove, PA, United States

J.T. den Dunnen Leiden University Medical Center, Leiden, Netherlands

E. Dequeker University of Leuven, Leuven, Belgium

M. De Rycke Universitair Ziekenhuis & Vrije Universiteit Brussel (VUB), Brussels, Belgium

A. Deshpande Los Alamos National Laboratory, Los Alamos, NM, United States

S. Drmanac Complete Genomics, Inc., Mountain View, CA, United States; BGI-Shenzhen, Shenzhen, China

R. Drmanac Complete Genomics, Inc., Mountain View, CA, United States; BGI-Shenzhen, Shenzhen, China

T. Ebai Uppsala University, Uppsala, Sweden

J.S. Farrar Virginia Commonwealth University School of Medicine, Richmond, VA, United States

L.R. Ferguson Nutrigenomics New Zealand, Auckland, New Zealand; The University of Auckland, Auckland, New Zealand

D.A. Forero Universidad Antonio Nariño, Bogotá, Colombia

J. Fredenburgh Scigen Scientific Inc., Gardena, CA, United States

B.K. Gale University of Utah, Salt Lake City, UT, United States

J. Göransson Uppsala University, Uppsala, Sweden

W.E. Grizzle University of Alabama at Birmingham, Birmingham, AL, United States

I. Grundberg Uppsala University, Uppsala, Sweden

I.G. Gut Barcelona Institute for Science and Technology, Barcelona, Spain

D. Heideman VU University Medical Center, Amsterdam, The Netherlands

I. Helbing Uppsala University, Uppsala, Sweden

S. Henriksson Uppsala University, Uppsala, Sweden

I. Hernández-Neuta Stockholm University, Solna, Sweden

F. Innocenti University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

M. Isaksson Uppsala University, Uppsala, Sweden

M. Jarvius Uppsala University, Uppsala, Sweden

H. Jayamohan University of Utah, Salt Lake City, UT, United States

M. Kamali-Moghaddam Uppsala University, Uppsala, Sweden

K. Kampourakis University of Geneva, Geneva, Switzerland

T. Katsila University of Patras School of Health Sciences, Patras, Greece

B. Koos Uppsala University, Uppsala, Sweden

L. Koumakis Foundation for Research and Technology – Hellas (FORTH), Heraklion, Greece

P. Laissue Del Rosario University, Bogotá, Colombia

U. Landegren Uppsala University, Uppsala, Sweden

C. Larsson Uppsala University, Uppsala, Sweden

K.M. Legg Center for Forensic Science Research and Education, Willow Grove, PA, United States

K.-J. Leuchowius Uppsala University, Uppsala, Sweden

H. Li University of Utah, Salt Lake City, UT, United States

J.S. Liu Complete Genomics, Inc., Mountain View, CA, United States

A. Llerena Extremadura University Hospital and Medical School, Badajoz, Spain

C. Lopez-Correa Genome Quebec, Montreal, QC, Canada

C. Mathieu KU Leuven, Leuven, Belgium

H.E. McKiernan Center for Forensic Science Research and Education, Willow Grove, PA, United States

E. Mendrinou University of Patras School of Health Sciences, Patras, Greece

A. Mezger Stockholm University, Solna, Sweden

K. Mitropoulos The Golden Helix Foundation, London, United Kingdom

C. Mizzi University of Malta, Msida, Malta

L. Moens Uppsala University, Uppsala, Sweden

Z. Mohamed University of Malaya, Kuala Lumpur, Malaysia

J. Nelson University of Utah, Salt Lake City, UT, United States

M. Nilsson Stockholm University, Solna, Sweden; Uppsala University, Uppsala, Sweden

L. Overbergh KU Leuven, Leuven, Belgium

A. Papachatzopoulou University of Patras School of Health Sciences, Patras, Greece

K. Pardali Uppsala University, Uppsala, Sweden

A.F. Patenaude Dana-Farber Cancer Institute, Boston, MA, United States

G.P. Patrinos University of Patras School of Health Sciences, Patras, Greece; Erasmus University Medical Center, Rotterdam, The Netherlands; United Arab Emirates University, Al-Ain, United Arab Emirates

P.C. Patsalis The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus

B.A. Peters Complete Genomics, Inc., Mountain View, CA, United States; BGI-Shenzhen, Shenzhen, China

G. Potamias Foundation for Research and Technology − Hellas (FORTH), Heraklion, Greece

N. Reisdorph University of Colorado Denver, Aurora, CO, United States

V. Romanov University of Utah, Salt Lake City, UT, United States

R. Samuel University of Utah, Salt Lake City, UT, United States

K.C. Sexton University of Alabama at Birmingham, Birmingham, AL, United States

A. Sgourou Hellenic Open University, Patras, Greece

D. Sie VU University Medical Center, Amsterdam, The Netherlands

E. Sistermans VU University Medical Center, Amsterdam, The Netherlands

O. Söderberg Uppsala University, Uppsala, Sweden

J. Son University of Utah, Salt Lake City, UT, United States

A. Squassina University of Cagliari, Cagliari, Italy

C. Staessen Universitair Ziekenhuis & Vrije Universiteit Brussel (VUB), Brussels, Belgium

J. Stenberg Uppsala University, Uppsala, Sweden

P.E.M. Taschner University of Applied Sciences Leiden, Leiden, Netherlands

H. Tönnies Robert Koch-Institute, Berlin, Germany

J. Tost CEA−Institut de Génomique, Evry, France

G. Tzimas Technological Educational Institute of Western Greece, Patras, Greece

V. Velissariou Bioiatriki Health Services, Athens, Greece

E. Viennas University of Patras School of Health Sciences, Patras, Greece

S. Vig KU Leuven, Leuven, Belgium

P.S. White Los Alamos National Laboratory, Los Alamos, NM, United States

C.T. Wittwer University of Utah School of Medicine, Salt Lake City, UT, United States

A. Wonkam University of Cape Town, Cape Town, South Africa

X. Xun Complete Genomics, Inc., Mountain View, CA, United States; BGI-Shenzhen, Shenzhen, China

B. Ylstra VU University Medical Center, Amsterdam, The Netherlands

Preface, Third Edition

We are delighted to deliver to the scientific community the third edition of *Molecular Diagnostics*, perhaps the most successful textbook in the field of molecular genetic testing, almost 12 years from its first appearance in the scientific literature.

In 2003, just 2 years after the publication of the first draft of the Human Genome sequence, we proposed to edit a textbook exclusively devoted to the description of molecular techniques used to identify the underlying genetic heterogeneity leading to inherited disorders. The first edition of *Molecular Diagnostics* was published in April 2005, followed by the second edition in October 2009. This textbook has been available to the scientific community for over 10 years now, and it is clearly considered to be the key reference in the field, judging from the following: (1) the large number of copies sold worldwide; (2) the various postgraduate and specialist training courses on Molecular Diagnostics that have been used as syllabus and course material; (3) the adoption from universities as the textbook for related undergraduate courses and curricula, which also led to its translation in 2008; and (4) the various positive reviews obtained not only from external reviewers in scientific journals and elsewhere but also from fellow academics and students.

This has prompted Elsevier/Academic Press to request the compilation of a third edition, justified not only from technological advances, particularly in high-throughput methods, but also from the intellectual revolution in biomedical sciences. In this third edition we decided to keep the original structure of the previous two editions, since this was one of its main innovative aspects, but we opted to expand the editorial team. We also decided to reshuffle the table of contents completely, providing a succinct outline and a historical perspective of the low-throughput methods that set a solid basis for the recent discoveries of the high-throughput methods together with a detailed overview of modern high-throughput methodologies, such as next-generation sequencing and microarray-based methods, along with examples of their applications in a modern molecular genetic testing laboratory.

The contents of this book are divided into three parts. The first part is dedicated to the battery of the most modern molecular biology techniques and a historical perspective of the low-throughput methods. In order to keep pace with recent developments, the majority of the chapters from the previous editions have been either merged into a few chapters or omitted altogether, while being mentioned and fully referenced in the updated chapters. A large number of chapters pertaining to high-throughput molecular diagnostic approaches, for example, microarrays, next-generation sequencing, mass spectrometry, next-generation sequencing cytogenomics, etc., have been included. The remaining chapters from the previous editions have been comprehensively updated to include not only technology innovations but also novel diagnostic applications. This resulted in the book being completely revamped with over half of its content being new compilations.

The second part attempts to integrate previously analyzed technology with different aspects of molecular diagnostics, such as pharmacogenomics, molecular forensics and victim identification in mass disasters, and preimplantation genetic diagnosis, while new emerging disciplines, such as nutrigenomics, genome informatics, and genomic databases, have been included. Finally, various everyday issues in a diagnostic laboratory, from genetic counseling to related ethical and psychological issues to safety and quality management, are discussed in the third and final part of the book. As with the previous two editions, we feel that the inclusion of the latter issues in this reference book has great relevance to our society.

As with the previous two editions, our effort has been assisted by many internationally renowned experts in their fields from five continents who kindly accepted our invitation to compile the 29 chapters of this book and share with us and our readers their expertise, experience, and results. In addition, we made an effort to formulate the book contents such that the notions described are explained in a simple language and terminology for the book to be useful not only to experienced physicians, healthcare specialists, and academics but also to undergraduate medical and life science students. The numerous self-explanatory illustrations and glossary clearly contribute to this end. Last, but not least, we provided the means to resolve the previously reported deficiencies in variant nomenclature by including a chapter on the official gene and genetic variation nomenclature at the very beginning of the textbook.

We are grateful to those colleagues who provided constructive comments and criticisms on the previous two editions and identified deficiencies that have been, hopefully, rectified in this third edition. However, we expect that some points in this book can still be further improved. Therefore we would again welcome comments and criticism from attentive readers, which will contribute to improving the contents of this book even further in its future editions. We are also grateful to the editors, Drs. Tari Broderick, Jeffrey Rossetti, and Tracy Tufaga at Elsevier, who helped us in close collaboration to overcome encountered difficulties. We also express our gratitude to all of the contributors for delivering outstanding compilations that summarize their experience and many years of hard work in their fields of research. We are indebted to Greg Harris, who was responsible for the design and the cover of this book, and to the production project managers Kirsty Halterman and Karen East, who have refined the final manuscript prior to going into production. We owe our thanks to the academic reviewers for their constructive criticisms on the chapters and their positive evaluation of our proposal for this third edition.

Last, but not least, we wish to cordially thank our families, from whom we have taken a considerable amount of time to devote to this project, for their patience and continuous support over the years.

Editor-in-Chief
George P. Patrinos
Department of Pharmacy, Faculty of Health Sciences
University of Patras, Patras, Greece
Associate Editors
Wilhelm J. Ansorge
École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
Phillip B. Danielson
Department of Biological Sciences, University of Denver
Denver, CO, United States of America
December 2016

Chapter 1

Molecular Diagnostics: Past, Present, and Future

G.P. Patrinos[1], P.B. Danielson[2,3] and W.J. Ansorge[4]
[1]University of Patras School of Health Sciences, Patras, Greece; [2]University of Denver, Denver, CO, United States; [3]Center for Forensic Science Research and Education, Willow Grove, PA, United States; [4]Ecole Polytechnique Federal Lausanne, EPFL, Lausanne, Switzerland

1.1 INTRODUCTION

Molecular (or nucleic acid-based) diagnosis of human disorders is referred to as the detection of genomic variants that are pathogenic and/or benign in DNA and/or RNA samples in order to facilitate detection, diagnosis, subclassification, prognosis, and monitoring response to therapy. Molecular diagnostics combines laboratory medicine with the knowledge and technology of molecular genetics and has been enormously revolutionized over the last decades, benefiting from discoveries in the fields of molecular biology and genomic technologies (Table 1.1). The identification and fine characterization of the genetic basis of inherited diseases is vital for the accurate provision of diagnosis. Gene discovery, via high-throughput methods, such as next-generation sequencing or genome-wide association studies, provides invaluable insights into the mechanisms of disease, and genomic markers allow physicians to not only assess disease predisposition but also to design and implement accurate diagnostic methods. The latter is of great importance, as the plethora and variety of molecular defects demands the use of multiple rather than a single variant detection platform. Molecular diagnostics has gradually become a clinical reality with its roots deep into the basic science of gene expression and gene function.

1.2 HISTORY OF MOLECULAR DIAGNOSTICS: INVENTING THE WHEEL

In 1949, Pauling and his coworkers introduced the term *molecular disease* in the medical vocabulary, based on their discovery that a single amino acid change at the b-globin chain leads to sickle cell anemia, characterized mainly by recurrent episodes of acute pain due to vessel occlusion. In principle, their findings set the foundations of molecular diagnostics, although the big revolution occurred many years later. At that time, when molecular biology was only hectically expanding, the provision of molecular diagnostic services was inconceivable and technically not feasible. The first seeds of molecular diagnostics were provided in the early days of recombinant DNA technology, with many scientists from various disciplines working in concert. cDNA cloning and sequencing were at that time invaluable tools for providing basic knowledge on the primary sequence of various genes. The latter provided a number of DNA probes, allowing for analysis via southern blotting of genomic regions, leading to the concept and application of restriction fragment length polymorphism (RFLP) to track a variant allele from heterozygous parents to a high-risk pregnancy. In 1976, Kan and coworkers carried out, for the first time, prenatal diagnosis of α-thalassemia, using hybridization on DNA isolated from fetal fibroblasts. Also, Kan and Dozy (1978), implemented RFLP analysis to pinpoint sickle cell alleles of African descent. This breakthrough provided the means of establishing similar diagnostic approaches for the characterization of other genetic diseases, such as phenylketonurea (Woo et al., 1983), cystic fibrosis (Farrall et al., 1986), and so on.

At that time, however, a significant technical bottleneck had to be overcome. The identification of the pathogenic variant was possible only through the construction of a genomic DNA library from the affected individual, in order to first clone the variant allele and then determine its nucleotide sequence. Again, many human globin gene mutations were among the first to be identified through such approaches (Busslinger et al., 1981; Treisman et al.,

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00001-8
Copyright © 2017 Elsevier Ltd. All rights reserved.

TABLE 1.1 Timeline of the Principal Discoveries in the Field of Molecular Biology, Which Influenced the Development of Molecular Diagnostics

Date	Discovery
1949	Characterization of sickle cell anemia as a molecular disease
1953	Discovery of the DNA double helix
1958	Isolation of DNA polymerases
1960	First hybridization techniques
1969	In situ hybridization
1970	Discovery of restriction enzymes and reverse transcriptase
1975	Southern blotting
1977	DNA sequencing
1983	First synthesis of oligonucleotides
1985	Restriction fragment length polymorphism analysis
1985	Invention of the polymerase chain reaction
1986	Development of fluorescent in situ hybridization
1988	Discovery of the thermostable DNA polymerase—optimization of the polymerase chain reaction
1992	Conception of the real-time polymerase chain reaction
1993	Discovery of structure-specific endonucleases for cleavage assays
1996	First application of DNA microarrays
2001	First draft versions of the human genome sequence
2001	Application of protein profiling in human diseases
2002	Launch of the HapMap project
2005	Introduction of high-throughput next-generation sequencing technology
2008	Launch of the 1000 Genomes Project
2013	Introduction of the CRISPR system for gene editing
2014	Announcement of the sequencing of the human genome for $1000
2015	Launch of the Precision Medicine Initiative by US President Barack Obama

1983). In 1982, Orkin and his coworkers showed that a number of sequence variations were linked to specific pathogenic *HBB* gene variants. These groups of RFLPs, termed *haplotypes* (both intergenic and intragenic), have provided a first-screening approach in order to detect a disease-causing variant. Although this approach enabled researchers to predict which *HBB* allele was pathogenic, significantly facilitating mutation screening, no one was in the position to determine the exact nature of the disease-causing mutation, as many different *HBB* gene variants were linked to a specific haplotype in different populations (further information is available at http://globin.bx.psu.edu/hbvar; Patrinos et al., 2004; Giardine et al., 2014).

At the same time, in order to provide a shortcut to DNA sequencing, a number of exploratory methods for pinpointing pathogenic variants in patients' DNA were developed, setting the basis for variant screening and scanning methods. The first methods involved mismatch detection in DNA/DNA or RNA/DNA heteroduplexes (Myers et al., 1985a,b) or differentiation of mismatched DNA heteroduplexes using gel electrophoresis, according to their melting profile (Myers et al., 1987). Using this laborious and time-consuming approach, a number of variant sequence alleles have been identified, which made possible the design of short synthetic oligonucleotides that were used as allele-specific probes onto genomic Southern blots. This experimental design was quickly implemented for the detection of β-thalassemia mutations (Orkin et al., 1983; Pirastu et al., 1983).

Despite intense efforts from different laboratories worldwide, the diagnosis of inherited diseases on the DNA level was still underdeveloped and therefore still not ready

to be implemented in clinical laboratories for routine analysis of patients due to the complexities, costs, and time requirements of the technology available. It was only after a few years that molecular diagnosis entered its golden era with the discovery of the most powerful molecular biology tool since cloning and sequencing, the polymerase chain reaction (PCR).

1.3 THE POST-POLYMERASE CHAIN REACTION REVOLUTION

The discovery of PCR (Saiki et al., 1985; Mullis and Faloona, 1987) and its quick optimization, using a thermostable *Taq* DNA polymerase from *Thermus aquaticus* (Saiki et al., 1988), has greatly facilitated and in principle revolutionized molecular diagnostics. The most powerful feature of PCR is the large amount of copies of the target sequence generated by its exponential amplification, which allows the identification of a known mutation within a single day, rather than months. Also, PCR has markedly decreased or even diminished the use of radioactivity for routine molecular diagnosis. This has allowed molecular diagnostics to enter the clinical laboratory for the provision of genetic services, such as carrier or population genetic screening, prenatal diagnosis of inherited diseases, or, in recent years, the identification of unknown variants, in close collaboration with research laboratories. Therefore by being moved to their proper environment, the clinical laboratory, molecular diagnostics could provide the services for which they have been initially conceived.

The discovery of PCR also has provided the foundations for the design and development of many variant detection schemes, based on amplified DNA. In general, PCR is either used for the generation of DNA fragments to be analyzed or is part of the detection method. The first attempt was the use of restriction enzymes (Saiki et al., 1985) or oligonucleotide probes, immobilized onto membranes or in solution (Saiki et al., 1986), in order to detect the existing genetic variation, in particular the sickle cell disease-causing mutation. In the following years, an even larger number of variant detection approaches have been developed and implemented (see also Chapter 3). These techniques can be divided roughly into three categories, depending on the basis for discriminating the allelic variants:

1. *Enzymatic-based methods*. RFLP analysis was historically the first widely used approach, exploiting the alterations in restriction enzyme sites, leading to the gain or loss of restriction events (Saiki et al., 1985). Subsequently, a number of enzymatic approaches for variant allele detection have been conceived, based on the dependence of a secondary structure on the primary DNA sequence. These methods exploit the activity of resolvase enzymes T4 endonuclease VII and T7 endonuclease I to digest heteroduplex DNA formed by annealing wild type and mutant DNA (Mashal et al., 1995). Digestion fragments indicate the presence and the position of any variants. A variation of the theme involves the use of chemical agents for the same purpose (Saleeba et al., 1992). Another enzymatic approach for variant detection is the oligonucleotide ligation assay (Landegren et al., 1988; Chapter 3); ligation was also one of the main principles of one of the most widely used next-generation sequencing approaches (see also Chapter 8).
2. *Electrophoretic-based techniques*. This category is characterized by a plethora of different approaches designed for the screening of known or unknown mutations, based on the different electrophoretic mobility of the mutant alleles, under denaturing or nondenaturing conditions. Single-strand conformation polymorphism (SSCP) and heteroduplex analyses (HDA; Orita et al., 1989; see Chapter 3) were among the first methods designed to detect molecular defects in genomic loci. In combination with capillary electrophoresis, SSCP and HDA analysis now provide an excellent, simple, and rapid variant detection platform with low operation costs and, most interestingly, the potential of easily being automated, thus allowing for high-throughput analysis of patients' DNA. Similarly, denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis can be used equally well for variant allele detection (see Chapter 3). In this case, electrophoretic mobility differences between a wild type and variant allele can be "visualized" in a gradient of denaturing agents, such as urea and formamide, or of increasing temperature. A less common variant detection technique is two-dimensional gene scanning, based on two-dimensional electrophoretic separation of amplified DNA fragments, according to their size and base pair sequence. The latter involves DGGE, following the size separation step.
3. *Solid phase-based techniques*. This set of techniques consists of the basis for most of the present-day mutation detection technologies, since they have the extra advantage of being easily automated and hence are highly recommended for high-throughput mutation detection or screening. A fast, accurate, and convenient method for the detection of known mutations is reverse dot-blot, initially developed by Saiki et al. (1989) and implemented for the detection of *HBB* gene variants leading to β-thalassemia. The essence of this method is the utilization of oligonucleotides, bound to a membrane, as hybridization targets for amplified DNA. Some of this technique's advantages are that one membrane strip can be used to detect many different known mutations in a single individual (a one strip-one patient

type of assay), the potential of automation, and the ease of interpretation of the results, using a classical avidin-biotin system. However, this technique cannot be used for the detection of unknown mutations. Continuous development has given rise to allele-specific hybridization of amplified DNA [PCR-ASO (Allele Specific Oligonucleotide), Chapter 3] on filters, recently extended to DNA oligonucleotide microarrays (see Chapter 18) for high-throughput mutation analysis (Gemignani et al., 2002; Cremonesi et al., 2007). In particular, oligonucleotides of known sequence are immobilized onto appropriate surfaces, and hybridization of the targets to the microarray is detected, mostly using fluorescent dyes.

The choice of the variant detection method is dependent upon a number of variables, including the variation spectrum of a given inherited disorder, the available infrastructure, the number of tests performed in the diagnostic laboratory, and issues of intellectual properties (see also Section 1.5.1). Most of the clinical diagnostic laboratories have not invested in expensive high-technology infrastructures, since the test volumes (the number of tests) expected to be performed have not been large enough to justify the capital investment. Therefore simple "home-brew"screening tests such as SSCP and HDA were and still are the methods of choice for many clinical laboratories, as they allow for rapid and simultaneous detection of different sequence variations at a detection rate of close to 100%. Although DNA amplification has significantly facilitated the expansion of molecular diagnostics, it nonetheless has a number of limitations, such as amplification of cytidine-guanine repeat-rich regions, the error-prone features of *Taq* polymerase (at a range of 10^{-4} to 10^{-5} per nucleotide), and so on. Finally, it is noteworthy that despite the wealth of variant detection methodologies, DNA sequencing, particularly in the era of whole genome sequencing and the breakthroughs in next-generation sequencing technology, is considered the golden standard and the definitive experimental procedure for variant calling. However, the costs for the initial investment and the difficulties for standardization and interpretation of ambiguous results have restricted its use especially to basic research laboratories.

1.4 MOLECULAR DIAGNOSTICS IN THE POST-GENOMIC ERA

In February 2001, with the announcement of the first draft sequence of the human genome (International Human Genome Sequencing Consortium, 2001; Venter et al., 2001) and subsequently with the genomic sequence of other organisms, molecular biology has entered into a new era with unprecedented opportunities and challenges. These tremendous developments put pressure on a variety of disciplines to intensify their research efforts to improve by orders of magnitude the existing methods for genomic variant detection, to make available data sets with genomic variation and analyze these sets using specialized software, to standardize and commercialize genetic tests for routine diagnosis, and to improve the existing technology in order to provide state-of-the-art automated devices for high-throughput genetic analysis.

The biggest challenge, following the publication of the human genome draft sequence, was to improve the existing variant detection technologies to achieve robust, cost-effective, rapid, and high-throughput analysis of genomic variation. Also, the increased pace of novel variant detection and gene discovery dictated the harmonization of gene nomenclature, an effort that was spearheaded by the Human Genome Variation Society (http://www.hgvs.org; see also Chapter 2). Since 2005, genomic technology has improved rapidly, and new high-throughput variant detection techniques have become available, whereas old methodologies have evolved to fit into the increasing demand for automated and high-throughput screening or were gradually abandoned. The chromatographic detection of polymorphic changes of pathogenic variants using denaturing high-performance liquid chromatography (DHPLC; for review, see Xiao and Oefner, 2001) is one of the new technologies that emerged. DHPLC reveals the presence of a genetic variation by the differential retention of homo- and heteroduplex DNA on reversed phase chromatography under partial denaturation. Single-base substitutions, deletions, and insertions can be detected successfully by ultraviolet or fluorescence monitoring within 2 to 3 min in unpurified PCR products as large as 1.5-kilo bases. These features, together with its low cost, make DHPLC one of the most powerful tools for mutational analysis. Also, pyrosequencing, a nongel-based genotyping technology, provides a very reliable method and an attractive alternative to DHPLC. Pyrosequencing detects de novo incorporation of nucleotides based on the specific template. The incorporation process releases a pyrophosphate, which is converted to ATP and followed by luciferase stimulation. The light produced, detected by a charge couple device camera, is "translated" to a pyrogram, from which the nucleotide sequence can be deducted (Ronaghi et al., 1998). This approach constituted the basis for the development of the first next-generation sequencing approaches by 454 Life Technologies (see also Chapter 8).

One of the major advances was the invention of the real-time PCR and the numerous variations of this theme (Holland et al., 1991; see Chapter 4). The method allows for the direct detection of the PCR product during the exponential (mid-log) phase of the reaction, therefore combining amplification and detection in one single step. The increased speed of real-time PCR is due largely to reduced cycles, the removal of post-PCR detection

procedures, and the use of fluorogenic labels and sensitive methods of detecting their emissions. Therefore real-time PCR is a very accurate and sensitive methodology with a variety of applications in molecular diagnostics, allows a high-throughput, and can easily be automated and performed on very small volumes, which makes it the method of choice for many modern diagnostic laboratories.

Above all, the DNA microarray-based genotyping approaches offer simultaneous analysis of many sequence alterations (see Chapter 18). In particular, microarrays consist of hundreds of thousands up to millions of oligonucleotides attached on a solid surface in an ordered array fashion. The DNA sample of interest is PCR-amplified and then hybridized onto the microarray. Each oligonucleotide in the high-density array acts as an allele-specific probe, and therefore perfectly matched sequences hybridize more efficiently to their corresponding oligonucleotides on the array. The hybridization signals are quantified by high-resolution fluorescent scanning and analyzed by computer software, resulting in the identification of the genotype in the corresponding places in the genome. Therefore using a high-density microarray makes possible the simultaneous detection of a great number of DNA alterations, hence facilitating genome-wide screening.

There has been a significant development of proteomics, which has the potential to become an indispensable tool for molecular diagnostics. A useful repertoire of proteomic technologies is available, with the potential to undergo significant technological improvements, which would be beneficial for increased sensitivity and throughput while reducing the sample requirement (see Chapter 13). The improvement of these technologies is a significant advance toward the need for better disease diagnostics. The detection of disease-specific protein profiles goes back to the use of two-dimensional protein gels (Hanash, 2000), when it was demonstrated that leukemias could be classified into different subtypes based on the different protein profile (Hanash et al., 2002). Nowadays, mass spectrometers are able to resolve many protein and peptide species in body fluids, being virtually set to revolutionize protein-based disease diagnostics (see Chapter 13). The robust and high-throughput nature of the mass spectrometric instrumentation is unparalleled and imminently suited for future clinical applications, as elegantly demonstrated by many retrospective studies in cancer patients (reviewed in Petricoin et al., 2002). Also, high-throughput protein microarrays, constructed from recombinant, purified, and yet functional proteins, allow the miniaturized and parallel analysis of large numbers of diagnostic markers in complex samples. The first pilot studies on disease tissues, such as assessing protein expression profiles in tissue derived from squamous cell carcinomas of the oral cavity (Knezevic et al., 2001) or the identification of proteins that induce an acute antibody response in autoimmune disorders, using auto-antigen arrays (Robinson et al., 2002), emerged in the early 2000s, indicating that proteomic pattern analysis ultimately might be applied as a screening tool for cancer in high-risk and general populations.

The development of state-of-the-art variant detection platforms not only has a positive impact on molecular genetic testing of inherited disorders but also provides technical means to other disciplines, such as to ascertain genetically modified (GM) products, which may contaminate non-GM seeds, or food ingredients containing additives and flavorings that have been genetically modified or have been produced from GM organisms, or the genotype of an animal strain. Molecular genetic testing is also applicable in the individualization of drug doses, also known as pharmacogenomics (Chapter 16), referred to as the delineation of interindividual genetic variability in genes that are mainly involved in drug metabolism and transport with drug efficacy and adverse effects. This approach amalgamates technological expertise from high-throughput omics approaches, such as genomics, transcriptomics, and functional genomics, to define and predict the nature of the response of an individual to a drug treatment and to rationally design newer drugs or improve existing ones (Squassina et al., 2010). The same can be also applicable in the personalization of diet, in an emerging discipline known as Nutrigenomics (Chapter 17). Ultimately, the identified genomic sequence variants need to be organized and stored in well-curated and specialized variant databases, enabling a physician or researcher to query and retrieve information relevant to diagnostic issues (see Chapter 20).

Finally, DNA analysis and testing have also significantly revolutionized the forensic sciences. Technical advances in molecular biology and increasing knowledge of the human genome have had a major impact on forensic medicine (see Chapter 21). Genetic characterization of individuals at the DNA level enables identity testing from a minimal amount of biological specimen, such as hair, blood, semen, bone, and so forth, in cases of sexual assault, homicide, and unknown human remains, and paternity testing is also changing from the level of gene products to the genomic level. DNA testing is by far more advantageous over conventional forensic methods and over the years has contributed to the acquittal of falsely accused people (saving most of them even from death row), the identification of individuals who committed criminal acts (Cohen, 1995), and even helped to specify identities of unknown human remains, such as those from the victims at Ground Zero in New York or from the skeletons of the Romanov family members (Gill et al., 1994; see Chapter 22).

1.5 FUTURE PERSPECTIVES: WHAT LIES BEYOND

Molecular diagnostics stem back to the spring of 1953, when the DNA double helix was first announced. Today, molecular diagnostics embody a set of high-throughput technological advances, which reveal the genotype in thousands of genomic positions and the entire genome at a very high accuracy and gradually decreasing costs. This also leads to a better understanding of the basis of inherited diseases, therefore allowing molecular diagnostics to play a key role in patient or disease management. Presently, a great number of samples are analyzed annually worldwide in both public and private laboratories, and the number of genetic tests available is steadily increased year by year, making molecular diagnostic laboratories indispensable in laboratory medicine. With the existing sequencing platforms, newborns can be screened for a number of inherited treatable diseases, while it is also possible that in the not-so-distant future, children at high risk for coronary artery disease will be identified and treated to prevent changes in their vascular walls during adulthood. Similarly, parents will have the option of being informed about their carrier status for many recessive diseases before they decide to start a family. Also, for middle-aged and older populations, scientists will be able to determine risk profiles for various late-onset diseases, preferably before the appearance of symptoms, which at least could be partly prevented through dietary (Chapter 17) or pharmaceutical interventions (Chapter 16). In the near future, the preemptive genotyping of genes involved in drug metabolism and transport will help toward individualizing drug response and will become indispensable in standard medical practice. Although the majority of these issues are gradually becoming a reality, some of them are still based on promises, though quite optimistic ones. Thus some of the new perspectives of the field could simply be a decade away.

1.5.1 Commercializing Molecular Diagnostics

Currently, clinical molecular genetics is part of mainstream health care worldwide with a molecular diagnostic unit or department within each health care unit. Although the notion of molecular diagnostics has increasingly gained momentum, genetic tests are still not generally used for population screening but rather for diagnosis, carrier screening, and prenatal diagnosis and only on a limited basis. The lack of cost-effectiveness analyses is one of the reasons (Snyder et al., 2014). Therefore in order to make molecular diagnostics widely available, several obstacles and issues need to be taken into consideration and resolved.

The first important issue is the choice of the technology and variant detection platform. Despite the fact that a plethora of variant detection methods is available, "… the best platform is the one that works best in your own laboratory." There is ample choice for genotyping, such as filters, gels, microarrays, and microtiter plates, for amplification-based technologies, for separation techniques, such as blotting, capillary electrophoresis, microarrays, and mass spectroscopy, and finally different means for labeling, such as radioactive, fluorescent, chemiluminescent, or enzymatic substances. The variety of detection approaches makes it difficult to determine which one is better suited for a laboratory setting. Generally speaking, DNA sequencing, especially in the next-generation sequencing era, is the golden standard for the identification of, causative or benign, DNA sequence variations (see also Chapter 9), but then again different issues have to be addressed such as the sequencing coverage, the enrichment methods, the genes to be analyzed, and so forth. The initial investment costs and the expected test volume are some of the factors that need to be taken into consideration prior to choosing the detection technique. Related issues are the costs of the adjacent hardware and software, testing reagents, and kits. The latter is of great importance, since most of the diagnostic laboratories today are running "home-brew" assays—for example, not using well-standardized genetic testing kits due to cost barriers, which brings to surface the issue of quality control of the reagents (see Chapter 29) and safety (see Chapter 28).

Another very important issue is training the personnel of a molecular diagnostic laboratory, reflected in the quality and the correct interpretation of results. Continuous genetic education of the personnel of the diagnostic laboratory is crucial for the accuracy of the results provided, particularly in recently emerged disciplines such as Pharmacogenomics (Reydon et al., 2012; Kampourakis et al., 2014). Many times, such as in the case of prenatal or preimplantation diagnosis, irrevocable decisions need to be made, most of the time based on a simple test result. As a result of continuous training and proficiency testing schemes, there has been a significant reduction of the number of incorrect genotypes diagnosed (http://www.eurogentest.org). In the United States, there is a voluntary biannual proficiency test for molecular diagnostic laboratories, while in Europe, the EuroGenTest European Network of Excellence (http://www.eurogentest.org) has been founded to promote quality in molecular genetic testing through the provision of external quality assessment (proficiency testing schemes) and the organization of best practice meetings and publication of guidelines. It is generally true that many geneticists and nongeneticist physicians would benefit from continuous genetics education regarding the appropriate use of molecular diagnostic tests, which is necessary to evaluate the method preanalytically and to interpret results.

Legal considerations and ethical concerns are also hurdles that need to be overcome. One issue is the

reimbursement of diagnosis costs. At present, very few tests are reimbursed by insurance companies, and the necessary regulatory and legal framework is far from complete. "Legalizing" molecular testing by the adoption of relevant regulations would probably result in an increase in test volumes, and at the same time it can pose an immense barrier to uncontrolled genetic testing. Similarly, the need to obtain informed consent from the patient to be analyzed is also of great importance and should be encouraged and facilitated by the diagnostic laboratory, particularly in the case of incidental findings that arise from next-generation sequencing approaches (Roche and Berg, 2015).

On the other hand, the issue of intellectual properties hampers the wide commercialization of molecular diagnostics. Almost all the clinically relevant genes have been patented, and the terms that the patent holders offer vary considerably. Among the difficulties are the limiting choice of variant detection platforms, the large loyalties for reagent use, and the exclusive sublicenses that many companies grant to clinical laboratories, leading to monopolies. Since one of the biggest challenges that the clinical laboratory is facing is patent and regulatory compliance, partnerships and collaborations may be envisaged in order to take technology licenses to the diagnostic laboratory that will subsequently develop, standardize, and distribute the assays. This will partly alleviate some of the intellectual properties issues. Finally, the issue of the medical genetics specialty is more urgent than ever. In the United States, medical genetics has been formally recognized as a medical specialty only within the past 15 years, and in Europe, medical genetics only recently has been formally recognized as a specialty (http://www.eshg.org). The implementation of this decision is still facing substantial difficulties (http://www.eshg.org/geneticseurope.htm), which will probably take years to bypass. With the completion of the Human Genome Project, genetics has become the driving force in medical research and is now poised for integration into medical practice. An increase in the medical genetics workforce, including geneticists and genetic counselors, will be necessary, while the same is true with the emergence of new disciplines, such as genome informatics, needed for genome results interpretation (Potamias et al., 2014; see also Chapter 19). After all, the Human Genome Project has made information of inestimable diagnostic and therapeutic importance available, and the medical profession now has the obligation to rise to both the opportunities and challenges that this wealth of genetic information presents.

1.5.2 Personalized Medicine

The term "personalized medicine" refers to the practice of medicine where patients receive the most appropriate medical treatment, fitting dosage, and combination of drugs based on their genetic background. Some of the reasons for many types of adverse drug reactions are already known and are often related to polymorphic gene alleles of drug-metabolizing enzymes (Squassina et al., 2010). The application of high-throughput genotyping tools for the identification and screening of single nucleotide polymorphisms can eventually lead to the determination of the unique molecular signature of an individual in a relatively short period of time, especially with the advent of whole genome sequencing and its application in pharmacogenomics (Mizzi et al., 2014; Katsila and Patrinos, 2015).

This way, individual drug responses can be predicted from predetermined genetic variances correlated with a drug effect. In other words, this will allow the physician to provide the patient with a selective drug prescription (see Chapter 16). Also, specialized information technology systems will allow for translation of this information into a clinically meaningful format, which clinicians will then use to evaluate and adjust the drug dose to each patient, hence allowing for a complete patient analysis and drug evaluation, especially in a preemptive manner (Lakiotaki et al., 2016). In addition to these efforts, there is a growing need to incorporate this increasingly complex body of knowledge to standard medical practice. Incorporating pharmacogenomics-related courses into the standard curriculum of medical schools can potentially ensure that the forthcoming generation of clinicians and researchers will be familiar with the latest developments in that field and will be capable of providing patients with the expected benefits of personalized medicine (Pisanu et al., 2014).

Similarly, nutrigenomics (or nutritional genomics) investigates the interactions between nutrition and an individual's genome and the consequent downstream effects on their phenotype with the aim to provide tailored nutritional advice or develop specialist food products (see Chapter 17). In other words, nutrigenomics recognizes that a specific dietary advice that can be beneficial for one individual may be inappropriate, or actually harmful, to another. Although comparable to pharmacogenomics, nutrigenomics is still considered to be an emerging science, contrary to pharmacogenomics that is considered to have "come of age" (Pavlidis et al., 2015).

However, there are growing concerns on the ethical aspects of personalized medicine. First of all, equality in medical care needs to be ensured, when genetics foretell clinicians which patients would be less likely to benefit from a particular drug treatment. Second, it will become increasingly vital to devise operational tools for the prevention of stigmatization and discrimination of different populations, in particular on ethnic grounds (van Ommen, 2002), and therefore every precaution should be taken to eliminate all lingering prejudice and bias associated with the study of human genetic variation. Other dilemmas include the right to deny an available treatment to specific

patient populations according to genetic-derived indications, as is currently the case with prenatal diagnosis.

1.5.3 Personal Genomics

The ultimate goal in health care will be the efficient integration of molecular diagnostics with therapeutics. With the advent of next-generation sequencing in 2005 (Margulies et al., 2005) and the avalanche of developments in this field since then (see Chapters 8 and 9), experts believe that reasonably soon, people will be able to have their own genomes sequenced for under $1000. As a matter of fact, there are already claims that the human genome can be sequenced for $1000 with a low depth of coverage. This is going to involve sequencing technology that is much cheaper and faster than today's machines, both commercially available and prototypes, and several efforts are currently under way, often encouraged by major funding. Previously, the entire DNA sequence of only a handful of individuals had been sequenced, such as Craig Venter (Levy et al., 2007), Jim Watson (Wheeler et al., 2008), and so on, while today there are thousands of genomes that have been sequenced, providing useful insights into the genetic basis of human diseases. Yet, several million genomes should be sequenced so that further insights can be gained on how to better individualize treatment modalities (Table 1.2). In the future, a person may appear at the clinic for treatment, "carrying" his or her entire genome at hand, or alternatively, nanotechnology could eventually enable DNA analysis with a portable DNA sequencing device. Similarly, a personal genomics health coach may be able to provide timely information regarding one's health, based on both artificial intelligence- and machine learning-aided algorithms and the wealth of genomics data available in the literature (see also Chapter 19). Overall, the provision of accurate personalized medicine services will comprise a complex interplay between (pharmaco)genomics, metabolomics, transcriptomics, and metagenomics, mediated by accurate genome informatics translation services (Fig. 1.1).

Even though the expectations are high and companies are using these new technologies to provide information to individuals to predict health and disease outcomes, even behavioral traits, it is generally premature to make promises for clinically useful information from genomic analyses. Next to that, there is an inherent danger of overestimating the usefulness of the various personalized genomic tests, particularly those that can be ordered directly by consumers, also known as direct-to-consumer (DTC) genetic tests (Magnus et al., 2009). Unlike other genetic analyses, these tests provide a sheer amount of genetic information, but their diagnostic or prognostic value remains uncertain because of the lack of information about the influence of environmental and other factors and the weak association for the vast majority of genetic loci with disease. Several studies and metaanalyses suggest that there is insufficient scientific evidence to conclude that genomic profiles are useful in measuring the genetic risk for common diseases or in developing personalized diet (Pavlidis et al., 2015; 2016) and lifestyle recommendations for disease prevention (Patrinos et al., 2013). It is puzzling how some companies in the DTC sector use their clients' genetic profiles to tailor individualized diets and lifestyle recommendations. Also, it is noteworthy that some of the companies that provide personal genomic tests have no physicians involved in ordering these tests, with the argument that "…patients deserve direct access to their health information without a physician intermediary" (http://www.nytimes.com/2008/06/26/business/26gene.html). In a few cases, these companies were banned by regulatory authorities from further selling their services. Overall, there are questions as to whether and how to regulate these tests, about the extent to which they provide (useful) medical information, and the risks from misinterpreting them.

TABLE 1.2 Impact of Human Genome Sequencing in Advancing Knowledge in Human Genetics and Personalized Medicine

Number of genomes	Impact
Thousands	• Identification of human genomic variations • Technology development for the complete and error-free sequencing of the human genome for patients and the general population
Millions	• Understanding of the molecular basis of human genetic diseases • Optimization of the genotype–phenotype correlation • Development of software tools for genotype–phenotype correlation in patients' genomic profiles

Billions	• Predictive and personalized medicine
	  ◦ Treatment individualization
	  ◦ Improvement of quality of life
	  ◦ Prevention of human diseases

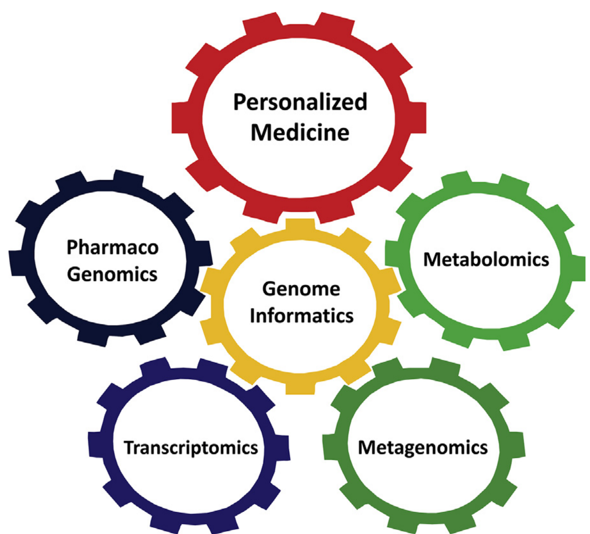


FIGURE 1.1 Schematic drawing depicting the interplay among the various omics disciplines that contribute to the implementation of personalized medicine.

1.6 CONCLUSIONS

In the coming years, molecular diagnostics will constitute an integral part of medical practice and public health worldwide. Molecular genetic testing will facilitate the detection and characterization of human disease as well as monitoring the drug response and will assist in the identification of genetic modifiers and disease susceptibility. Whole genome sequencing is likely to become the gold standard in assessing DNA variation and changes in gene expression. However, there are major hurdles to overcome before the implementation of these tests in clinical laboratories, such as which test to employ, the choice of technology and equipment, and issues such as cost-effectiveness, accuracy, reproducibility, personnel training, reimbursement by third-party payers, and intellectual property. Together with proteomic-based testing, these advances will improve molecular diagnostics and will present additional challenges for implementing such technology in public or private research units, hospitals, clinics, and pharmaceutical companies.

REFERENCES

Busslinger, M., Moschonas, N., Flavell, R.A., 1981. Beta + thalassemia: aberrant splicing results from a single point mutation in an intron. Cell 27, 289−298.

Cohen, J., 1995. Genes and behavior make an appearance in the O. J. Trial. Sci. 268, 22−23.

Cremonesi, L., Ferrari, M., Giordano, P., Harteveld, C.L., Kleanthous, M., Papasavva, T., Patrinos, G.P., Traeger-Synodinos, J., 2007. An overview of current microarray-based human globin gene mutation detection methods. Hemoglobin 31, 289−311.

Farrall, M., Rodeck, C.H., Stanier, P., Lissens, W., Watson, E., Law, H.Y., Warren, R., Super, M., Scambler, P., Wainwright, B., Williamson, R., 1986. First-trimester prenatal diagnosis of cystic fibrosis with linked DNA probes. Lancet 327, 1402−1405.

Gemignani, F., Perra, C., Landi, S., Canzian, F., Kurg, A., Tonisson, N., Galanello, R., Cao, A., Metspalu, A., Romeo, G., 2002. Reliable detection of beta-thalassemia and G6PD mutations by a DNA microarray. Clin. Chem. 2002 (48), 2051−2054.

Giardine, B., Borg, J., Viennas, E., Pavlidis, C., Moradkhani, K., Joly, P., Bartsakoulia, M., Riemer, C., Miller, W., Tzimas, G., Wajcman, H., Hardison, R.C., Patrinos, G.P., 2014. Updates of the HbVar database of human hemoglobin variants and thalassemia mutations. Nucleic Acids Res. 42, D1063−D1069.

Gill, P., Ivanov, P.L., Kimpton, C., Piercy, R., Benson, N., Tully, G., Evett, I., Hagelberg, E., Sullivan, K., 1994. Identification of the remains of the Romanov family by DNA analysis. Nat. Genet. 6, 130−135.

Hanash, S.M., 2000. Biomedical applications of two-dimensional electrophoresis using immobilized pH gradients: current status. Electrophoresis 21, 1202−1209.

Hanash, S.M., Madoz-Gurpide, J., Misek, D.E., 2002. Identification of novel targets for cancer therapy using expression proteomics. Leukemia 16, 478−485.

Holland, P.M., Abramson, R.D., Watson, R., Gelfand, D.H., 1991. Detection of specific polymerase chain reaction product by utilizing the 5′−3′ exonuclease activity of Thermus aquaticus. Proc. Natl. Acad. Sci. U.S.A. 88, 7276−7280.

International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. Nature 409, 860−921.

Kampourakis, K., Vayena, E., Mitropoulou, C., Borg, J., van Schaik, R.H., Cooper, D.N., Patrinos, G.P., 2014. Key challenges for next generation pharmacogenomics. EMBO Rep. 15, 472−476.

Kan, Y.W., Dozy, A.M., 1978. Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation. Proc. Natl. Acad. Sci. U.S.A. 75, 5631−5635.

Kan, Y.W., Golbus, M.S., Dozy, A.M., 1976. Prenatal diagnosis of alpha-thalassemia. Clinical application of molecular hybridization. N. Engl. J. Med. 295, 1165−1167.

Katsila, T., Patrinos, G.P., 2015. Whole genome sequencing in pharmacogenomics. Front. Pharmacol. 6, 61.

Knezevic, V., Leethanakul, C., Bichsel, V.E., Worth, J.M., Prabhu, V.V., Gutkind, J.S., Liotta, L.A., Munson, P.J., Petricoin 3rd, E.F., Krizman, D.B., 2001. Proteomic profiling of the cancer microenvironment by antibody arrays. Proteomics 1, 1271−1278.

Lakiotaki, K., Kartsaki, E., Kanterakis, A., Katsila, T., Patrinos, G.P., Potamias, G., 2016. ePGA: a web-based information system for translational pharmacogenomics. PLoS One. http://dx.doi.org/10.1371/journal.pone.0162801 (in press).

Landegren, U., Kaiser, R., Sanders, J., Hood, L., 1988. A ligase-mediated gene detection technique. Science 241, 1077−1080.

Levy, S., Sutton, G., Ng, P.C., Feuk, L., Halpern, A.L., Walenz, B.P., Axelrod, N., Huang, J., Kirkness, E.F., Denisov, G., Lin, Y., MacDonald, J.R., Pang, A.W., Shago, M., Stockwell, T.B., Tsiamouri, A., Bafna, V., Bansal, V., Kravitz, S.A., Busam, D.A., Beeson, K.Y., McIntosh, T.C., Remington, K.A., Abril, J.F., Gill, J., Borman, J., Rogers, Y.H., Frazier, M.E., Scherer, S.W.,

Strausberg, R.L., Venter, J.C., 2007. The diploid genome sequence of an individual human. PLoS Biol. 5, e254.

Magnus, D., Cho, M.K., Cook-Deegan, R., 2009. Direct-to-consumer genetic tests: beyond medical regulation? Genome Med. 1, 17.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L., Jarvie, T.P., Jirage, K.B., Kim, J.B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., Rothberg, J.M., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376−380.

Mashal, R.D., Koontz, J., Sklar, J., 1995. Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. Nat. Genet. 9, 177−183.

Mizzi, C., Mitropoulou, C., Mitropoulos, K., Peters, B., Agarwal, M.R., van Schaik, R.H., Drmanac, R., Borg, J., Patrinos, G.P., 2014. Personalized pharmacogenomics profiling using whole genome sequencing. Pharmacogenomics 15, 1223−1234.

Mullis, K.B., Faloona, F.A., 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol. 155, 335−350.

Myers, R.M., Larin, Z., Maniatis, T., 1985a. Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA: DNA duplexes. Science 230, 1242−1246.

Myers, R.M., Lumelsky, N., Lerman, L.S., Maniatis, T., 1985b. Detection of single base substitutions in total genomic DNA. Nature 1985 (313), 495−498.

Myers, R.M., Maniatis, T., Lerman, L.S., 1987. Detection and localization of single base changes by denaturing gradient gel electrophoresis. Methods Enzymol. 155, 501−527.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., Sekiya, T., 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc. Natl. Acad. Sci. U.S.A. 86, 2766−2770.

Orkin, S.H., Kazazian Jr., H.H., Antonarakis, S.E., Goff, S.C., Boehm, C.D., Sexton, J.P., Waber, P.G., Giardina, P.J., 1982. Linkage of beta-thalassaemia mutations and beta-globin gene polymorphisms with DNA polymorphisms in human betaglobin gene cluster. Nature 296, 627−631.

Orkin, S.H., Markham, A.F., Kazazian Jr., H.H., 1983. Direct detection of the common Mediterranean beta-thalassemia gene with synthetic DNA probes. An alternative approach for prenatal diagnosis. J. Clin. Invest. 71, 775−779.

Patrinos, G.P., Baker, D.J., Al-Mulla, F., Vasiliou, V., Cooper, D.N., 2013. Genetic tests obtainable through pharmacies: the good, the bad and the ugly. Hum. Genomics 7 (1), 17.

Patrinos, G.P., Giardine, B., Riemer, C., Miller, W., Chui, D.H., Anagnou, N.P., Wajcman, H., Hardison, R.C., 2004. Improvements in the HbVar database of human hemoglobin variants and thalassemia mutations for population and sequence variation studies. Nucleic Acids Res. 32, D537−D541.

Pauling, L., Itano, H.A., Singer, S.J., Wells, I.C., 1949. Sickle cell anemia, a molecular disease. Science 110, 543−548.

Pavlidis, C., Lanara, Z., Balasopoulou, A., Nebel, J.C., Katsila, T., Patrinos, G.P., 2015a. Meta-analysis of nutrigenomic biomarkers denotes lack of association with dietary intake and nutrient-related pathologies. OMICS 19, 512−520.

Pavlidis, C., Patrinos, G.P., Katsila, T., 2015b. Nutrigenomics: a controversy. Appl. Transl. Genomics 4, 50−53.

Pavlidis, C., Nebel, J.C., Katsila, T., Patrinos, G.P., 2016. Nutrigenomics 2.0: the need for ongoing and independent evaluation and synthesis of commercial nutrigenomics tests' scientific knowledge base for responsible innovation. OMICS 20, 65−68.

Petricoin, E.F., Zoon, K.C., Kohn, E.C., Barrett, J.C., Liotta, L.A., 2002. Clinical proteomics: translating benchside promise into bedside reality. Nat. Rev. Drug. Discov. 1, 683−695.

Pirastu, M., Kan, Y.W., Cao, A., Conner, B.J., Teplitz, R.L., Wallace, R.B., 1983. Prenatal diagnosis of beta-thalassemia. Detection of a single nucleotide mutation in DNA. N. Engl. J. Med. 309, 284−287.

Pisanu, C., Tsermpini, E.E., Mavroidi, E., Katsila, T., Patrinos, G.P., Squassina, A., 2014. Assessment of the pharmacogenomics educational environment in Southeast Europe. Public Health Genomics 17, 272−279.

Potamias, G., Lakiotaki, K., Katsila, T., Lee, M., Topouzis, S., Cooper, D.N., Patrinos, G.P., 2014. Deciphering next-generation pharmacogenomics: an information technology perspective. Open Biol. 4, 140071.

Reydon, T.A., Kampourakis, K., Patrinos, G.P., 2012. Genetics, genomics and society: the responsibilities of scientists for science communication and education. Per. Med. 9, 633−643.

Robinson, W.H., DiGennaro, C., Hueber, W., Haab, B.B., Kamachi, M., Dean, E.J., Fournel, S., Fong, D., Genovese, M.C., de Vegvar, H.E., Skriner, K., Hirschberg, D.L., Morris, R.I., Muller, S., Pruijn, G.J., van Venrooij, W.J., Smolen, J.S., Brown, P.O., Steinman, L., Utz, P.J., 2002. Autoantigen microarrays for multiplex characterization of autoantibody responses. Nat. Med. 8, 295−301.

Roche, M.I., Berg, J.S., 2015. Incidental findings with genomic testing: implications for genetic counseling practice. Curr. Genet. Med. Rep. 3 (4), 166−176.

Ronaghi, M., Uhlen, M., Nyren, P., 1998. A sequencing method based on real-time pyrophosphate. Science 281, 363−365.

Saiki, R.K., Bugawan, T.L., Horn, G.T., Mullis, K.B., Erlich, H.A., 1986. Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. Nature 324, 163−166.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., Erlich, H.A., 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239, 487−491.

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., Arnheim, N., 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230, 1350−1354.

Saiki, R.K., Walsh, P.S., Levenson, C.H., Erlich, H.A., 1989. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. Proc. Natl. Acad. Sci. U.S.A. 86, 6230−6234.

Saleeba, J.A., Ramus, S.J., Cotton, R.G., 1992. Complete mutation detection using unlabeled chemical cleavage. Hum. Mutat. 1, 63−69.

Snyder, S.R., Mitropoulou, C., Patrinos, G.P., Williams, M.S., 2014. Economic evaluation of pharmacogenomics: a value-based approach to pragmatic decision making in the face of complexity. Public Health Genomics 17, 256−264.

Squassina, A., Artac, M., Manolopoulos, V.G., Karkabouna, S., Lappa-Manakou, C., Mitropoulos, K., Manchia, M., del Zompo, M., Patrinos, G.P., 2010. Translation of genetic knowledge into clinical practice: the expectations and realities of pharmacogenomics and personalized medicine. Pharmacogenomics 11, 1149−1167.

Treisman, R., Orkin, S.H., Maniatis, T., 1983. Specific transcription and RNA splicing defects in five cloned beta-thalassaemia genes. Nature 302, 591−596.

van Ommen, G.J., 2002. The Human genome project and the future of diagnostics, treatment and prevention. J. Inherit. Metab. Dis. 25, 183−188.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigo, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., Zhu, X., 2001. The sequence of the human genome. Science 291, 1304−1351.

Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.J., Makhijani, V., Roth, G.T., Gomes, X., Tartaro, K., Niazi, F., Turcotte, C.L., Irzyk, G.P., Lupski, J.R., Chinault, C., Song, X.Z., Liu, Y., Yuan, Y., Nazareth, L., Qin, X., Muzny, D.M., Margulies, M., Weinstock, G.M., Gibbs, R.A., Rothberg, J.M., 2008. The complete genome of an individual by massively parallel DNA sequencing. Nature 452, 872−876.

Woo, S.L., Lidsky, A.S., Guttler, F., Chandra, T., Robson, K.J., 1983. Cloned human phenylalanine hydroxylase gene allows prenatal diagnosis and carrier detection of classical phenylketonuria. Nature 306, 151−155.

Xiao, W., Oefner, P.J., 2001. Denaturing high-performance liquid chromatography: a review. Hum. Mutat. 17, 439−474.

Chapter 2

Describing DNA Variants (Mutation Nomenclature)

J.T. den Dunnen[1] **and P.E.M. Taschner**[2]

[1]Leiden University Medical Center, Leiden, Netherlands; [2]University of Applied Sciences Leiden, Leiden, Netherlands

2.1 INTRODUCTION

In DNA-based molecular diagnostics the focus is on trying to find one or more changes in the DNA sequence that can be causally linked to the health status of the individual analyzed. The variants detected need to be reported, i.e., described in a specific format. The current standard for reporting is the Human Genome Variation Society (HGVS) Recommendations for the Description of Sequence Variants. This chapter will describe the basic principles of the HGVS variant nomenclature covering different types of variants (substitution, deletion, duplication, insertion, etc.). In addition, examples will be given for the description of DNA changes as well as their consequences for corresponding RNA and protein sequences.

2.2 HISTORY

A first suggestion for the standardized description of sequence variants was made in 1993 (Beadet and Tsui, 1993). A major obstacle at the time was the scarcity of high-quality DNA sequences, making nucleotide numbering a topic of discussion. Preferably variant nomenclature should be accurate, unambiguous, and stable yet sufficiently flexible to allow for the description of all known classes of sequence variation. How could this be achieved when the availability of high-quality DNA sequences was very limited? Based on lively discussions many opinions evolved, the recommendations changed, and updates were published in 1996, 1998, and 2000 (Ad Hoc Committee on Mutation Nomenclature, 1996; Antonarakis, 1998; den Dunnen and Antonarakis, 2000). The latter recommendations, published on behalf of the HGVS, were promoted widely, e.g., made mandatory to use for publication in the journal *Human Mutation* (Wiley). These so-called HGVS recommendations are now considered to be a standard worldwide, used by many journals and recommended for use in molecular diagnostics (Gulley et al., 2007; Richards et al., 2015). The Human Variome Project (HVP) maintains a list of journals obliging or recommending the use of HGVS nomenclature and/or variant submissions to databases at http://www.humanvariomeproject.org/resources/genetics-and-genomics-journals.html.

2.3 AUTHORIZATION

The recommendations are currently commissioned under the auspices of three international organizations: HGVS, HVP, and the Human Genome Organization (HUGO). Requests for modifications and extensions go through the Sequence Variant Description Working Group (SVD-WG) operating following a standard procedure, including a final community consultation step. Version numbers are assigned to the recommendations, facilitating users to specify which version is used in their variant descriptions. The recommendations that will be described in this chapter follow HGVS version 15.11 (2015, November).

The HGVS nomenclature page lists an email address for questions (VarNomen@HGVS.org). Questions are either answered directly or after consultation with the SVD-WG. Facebook (http://www.facebook.com/HGVSmutnomen) is used to discuss topics of interest, including simplified Q&As, mentioning meetings where the recommendations are discussed, and the release of new proposals for community consultation. Basic educational material is available from the HGVS (a slide presentation entitled "The basics" and an online Q&A test); more is in preparation.

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00002-X
Copyright © 2017 Elsevier Ltd. All rights reserved.

2.4 DEFINITIONS

2.4.1 Principles

To prevent confusion, the recommendations avoid the use of ambiguous terms like "polymorphism" and "mutation." Polymorphism refers to "a sequence variant" in some disciplines and to "a variant found at a frequency of 1% or higher in a population" in other disciplines. Similarly, mutation is avoided since it is used to indicate both a "change" and a "disease-causing change." HGVS therefore only uses neutral terms like "variant," "alteration," and "change."

To facilitate the computational analysis and description of sequence variants, the basic types of variants have been precisely defined (see the following) and prioritized as (1) deletion, (2) inversion, (3) duplication, (4) conversion, and (5) insertion.

- **Substitution:** a change in a specific sequence where, compared to the reference sequence, ***one*** nucleotide is replaced by ***one*** other nucleotide
- **Deletion:** a change in a specific sequence where, compared to the reference sequence, one or more nucleotides are not present (deleted)
- **Duplication:** a change in a specific sequence where, compared to the reference sequence, a copy of one or more nucleotides is inserted ***directly 3′*** of the original copy of that sequence
- **Insertion:** a change in a specific sequence where, compared to the reference sequence, one or more nucleotides are inserted in a sequence and where the insertion is not a copy of a sequence immediately 5′ from the insertion location
- **Deletion-Insertion** (indel): a change in a specific sequence where, compared to the reference sequence, one or more nucleotides are replaced by one or more other nucleotides and which is not a substitution, inversion, or conversion
- **Inversion:** a change in a specific sequence where, compared to the reference sequence, ***more than one*** nucleotide replaces the original sequence and where this sequence is the ***reverse-complement*** of the original sequence (e.g., CTCGA to TCGAG)
- **Conversion** (a specific type of deletion-insertion): a change where a range of nucleotides from the original reference sequence is replaced by a range of nucleotides copied from a homologous sequence present at another site in the genome
- **Translocation:** a change in a specific sequence where, compared to the reference sequence, the sequence of one chromosome is joined to that of another chromosome

Following these definitions an A>T substitution cannot be described as an inversion, and a change where one nucleotide is replaced by more than one nucleotide cannot be described as a substitution (it is a deletion-insertion (indel)).

The prioritization determines the preferred class when a description is possible according to several classes, e.g., the addition of one or more nucleotides to an existing mononucleotide stretch is described as a duplication, not an insertion. Together definitions and prioritization made it possible to generate a formalized description of the HGVS standard in Extended Backus-Naur Form (Laros et al., 2011) and to develop software tools that can check and/or generate HGVS descriptions (Wildeman et al., 2008; Hart et al., 2014).

2.4.2 Reference Sequences

All variants have to be described in relation to a reference sequence (see Table 2.1) and the sample analyzed. When a variant is detected in a cDNA (RNA), it should be described using an RNA reference sequence (r.), and it can only be described using a DNA reference sequence when genomic DNA is analyzed.

TABLE 2.1 Reference Sequences

Numbering Scheme	Prefix	Position Numbering in Relation to
Genomic DNA	g.	First nucleotide of the genomic reference sequence
Coding DNA	c.	First nucleotide of the translation start codon of the coding DNA reference sequence
Noncoding DNA	n.	First nucleotide of the noncoding DNA reference sequence
Mitochondrial DNA	m.	First nucleotide of the mitochondrial DNA reference sequence
RNA	r.	First nucleotide of the translation start codon of the coding RNA reference sequence **or** first nucleotide of the noncoding RNA reference sequence
Protein	p.	First amino acid of the protein sequence

For diagnostic applications it is strongly recommended to use a Locus Reference Genomic sequence (LRG) (Dalgleish et al., 2010; MacArthur et al., 2014). When no LRG is available for a gene of interest, one should be requested as soon as possible.

The preferred reference sequence is a complete genomic reference sequence (i.e., the recent genome build GRCh38). In diagnostics, a genomic reference sequence with coding DNA annotation is preferred, since descriptions are usually shorter/simpler, and one can derive the location of the variant from the description (i.e., exonic/intronic, 5′/3′ of the ATG, the exonic position divided by 3 is the amino acid residue affected; see Table 2.3).

2.4.3 Locus Reference Genomic Sequence

The recommended reference sequence for diagnostic reporting is a Locus Reference Genomic sequence (LRG) (Dalgleish et al., 2010; MacArthur et al., 2014). The LRG has been developed to create a stable and reliable genomic reference sequence for diagnostic purposes. LRGs are not versioned, and their sequences are not changed during their lifetime. LRGs are annotated with transcript and protein information divided into two sections: a stable section and an updateable section. The HGVS recommendations strongly advise the use of an LRG. When an LRG for the gene analyzed is not available, one should be requested (http://www.lrg-sequence.org/lrg-request).

2.4.4 Numbering Residues

The numbering of residues (nucleotides or amino acids) starts at 1 for all reference sequences, except for the coding DNA reference sequence scheme. In a coding DNA reference sequence scheme, nucleotide numbering starts with 1 at the A of the ATG translation initiation codon and ends with the last nucleotide of the translation stop codon (Table 2.2). Nucleotides 5′ of the ATG are numbered … -3 -2 -1, and nucleotides 3′ of the stop codon *1 *2 *3 …. There is no nucleotide with the number 0. Intronic nucleotides are numbered relative to the closest coding DNA nucleotide, beginning at the 5′ end of the intron with +1 (e.g., c.77+1, c.77+2, c.77+3 …) and ending at the 3′ end with -1 (e.g., c78-3, c.78-2, c.78-1). The number prefix changes in the middle of the intron from "+" to "-". For introns with an uneven number of nucleotides, the central nucleotide is the last to be numbered with a "+".

For variants in stretches of repeated sequences (e.g., TTT, agcagcagc, HisHisHisHis) the so-called ***3′ rule*** applies, stating that the ***most 3′ position possible*** is arbitrarily deemed to have been changed. Consequently the change of GGG to GG is described as g.3del (not g.1del or g.2del). This rule applies to DNA, RNA, and protein sequences.

2.4.5 Coding Residues

At the DNA, RNA, and protein level, nucleotides and amino acids are described using the International Union of Pure and Applied Chemistry - International Union of Biochemistry and Molecular Biology (IUPAC-IUB) standards (see Table 2.3, IUPAC-IUB, 1984). For DNA, uppercase letters are used (A, C, G, T), and for RNA, lowercase letters are used (a, c, g, u). At the protein level the 3-letter amino acid code is preferred, and the translation stop codon is described using "Ter" or "*". Initial recommendations suggested the use of "X" to indicate the translation stop codon (Antonarakis, 1998; den Dunnen and

TABLE 2.2 Numbering Residues

	Genomic Reference Sequence		Coding DNA Reference Sequence
Location Nucleotide	**NC_000023.10**	**LRG_199 (DMD)**	**LRG_199t1, NM_004006.2**
5′ transcription start	g.33231774	g.130953	c.-2345 *(c.-244-u2101)*[c]
In 5′ UTR	g.33229552	g.133175	c.-123
(In intron in 5′ UTR)[a]			(c.-55+23/c.-54-23)[b]
A of the ATG start codon	g.33229429	g.133298	c.1
In coding DNA	g.32862930	g.499797	c.234
In intron, 5′ side	g.32380903	g.981883	c.5325+2[a]
In intron, 3′ side	g.32366647	g.996080	c.5326-2[a]
G of the TAG stop codon	g.31140036	g.2222791	c.11058
In 3′ UTR	g.31139691	g.2222836	c.*345
(In intron in 3′ UTR)[a]			(c.*54+23/c.*55-23)[b]
3′ transcription end	g.31136580	g.2226247	c.*3456 *(c.*2691+d756)*[c]

Nucleotide numbering using a genomic reference sequence (NC_000023.10 [genome build GRCh37/hg19] and LRGt1_199) and a coding DNA reference sequence (DMD gene, NM_004006.2).
[a]*Coding DNA reference sequence NM_004006.2 does not contain intron sequences; LRG_199t1 is required for this description.*
[b]*Hypothetical example: the DMD gene does not contain an intron in the 5′ or 3′ untranslated region (UTR).*
[c]*Proposed (not yet approved) to allow the discrimination of nucleotides that are 5′ (u = upstream) or 3′ (d = downstream) of the transcript (RNA).*

TABLE 2.3 Symbols for DNA (RNA)

Symbol	Meaning	Description
A	A	Adenine
C	C	Cytosine
G	G	Guanine
T	T	Thymine
B	C, G, or T	Not-A (B follows A in the alphabet)
D	A, G, or T	Not-C (D follows C in the alphabet)
H	A, C, or T	Not-G (H follows G in the alphabet)
K	G or T	Keto
M	A or C	aMino
N	A, C, G, or T	aNy
R	A or G	puRine
S	G or C	Strong interaction (3 H-bonds)
V	A, C, or G	Not-T/not-U (V follows U in the alphabet)
W	A or T	Weak interaction (2 H-bonds)
Y	C or T	pYrimidine

Symbols follow the International Union of Pure and Applied Chemistry recommendations (IUPAC-IUB, 1984). For RNA symbols change DNA symbols to lowercase and use “u” for “T”.

Antonarakis, 2000), but this recommendation has been retracted and changed to “Ter” or “*”.

2.5 VARIANT DESCRIPTIONS

The most fundamental recommendation is that one has to describe what ***was observed***, not what was ***inferred***, i.e., report original observations using an appropriate reference sequence. Descriptions at the protein level should make clear whether experimental proof was available or not. When not, one should list the predicted consequences in parentheses, e.g., in the format *p.(Leu83Cys)*.

2.5.1 Substitutions

The format to describe substitutions differs at the DNA/RNA level compared to the protein level. At the DNA/RNA level, the “>” character is used: *g.76198571C>T* (NC_000001.10), *c.250C>T* (NM_000016.4), *r.250c>u* (NM_000016.4). At the protein level, the “>” character is not used to describe substitution. The format used is *p.Leu84Phe* (NM_000016.4, NP_000007.1).

2.5.2 Deletions

Deletions are described using “del” after the position of the first and last residues deleted, separated by a “_” (underscore). For all descriptions, the most 3′ possible position is arbitrarily deemed to have been changed.

One residue: *g.215,914,514del* (NC_000,002.11), *c.529del* (NM_173,076.2), *r.529del* (NM_173,076.2), *p.(Pro2141del)* (NM_173,076.2, NP_775,099.2).

Several residues: *g.27,106,792_27,106,797del* (NC_000,001.10), *c.6403_6408del* (NM_006,015.4), *r.6403_6408del* (NM_006,015.4), *p.Ile2135_Leu2136del* (NM_006,015.4, NP_006,006.3).

When deletion breakpoints have not been sequenced they should be indicated as precisely as possible, describing the range of uncertainty in parentheses and using a question mark (“?”) to indicate unknown positions: (last-residue-present_first-residue-deleted)_(last-residue-deleted_first-residue-present), e.g., *g.(1234_3456)_(5678_7890)del* with defined breakpoints; *g.(?_3456)_(5678_?)del* with 5′ and 3′ undefined breakpoints; and *g.2345_(5678_7890)del* with one sequenced and one undefined breakpoint. Descriptions should not use exon or intron numbers (EX3_EX5del, IVS2_IVS5del).

2.5.3 Duplications

Duplications are described using “dup” after the position of the first and last residues duplicated, separated by a “_” (underscore). For all descriptions, the most 3′ possible position is arbitrarily deemed to have been changed.

One residue: *g.103,396,826dup* (NC_000,014.8), *c.1253dup* (NM_030,943.3), *r.1253dup* (NM_030,943.3), *p.(Leu46dup)* (NM_001,077,488.2, NP_001,070,956.1).

Several residues: *g.54,209,303_54,209,308dup* (NC_000,023.10), *c.324_329dup* (NM_017,848.4), *r.324_329dup* (NM_017,848.4), *p.Glu75_Asp76dup* (NM_017,848.4, NP_060,318.3).

2.5.4 Insertions

Insertions are described using “ins” after the position of the residues flanking the insertion site, followed by a description of the residues inserted. Duplicating insertions should be described as duplications, not insertions.

One residue: *g.103,396,826_103,396,827insT* (NC_000,014.8), *c.1253_1254insC* (NM_030,943.3), *r.1253_1254insc* (NM_030,943.3), *p.(Leu46_Gly46insArg)* (NM_001,077,488.2, NP_001,070,956.1).

Several residues: *g.54,209,303_54,209,304insTTAAC* (NC_000,023.10), *g.54,209,303_54,209,304ins78818454_78,819,003* (NC_000,023.10), c.*324_325insGAAGT* (NM_017,848.4), *324_325insGAAGT* (NM_017,848.4),

p.Val118_Glu119insArgCysThr (NM_017,848.4, NP_060,318.3).

Description like "ins8" are not allowed, since the insertion is not specified. Large insertions can be described by referring to a sequence file containing the complete inserted sequence with specification of accession.version numbers, e.g., NC_000,023.10:*g.54,209,303_54,209,304 insNC_000,006.10:78,818,454_78,819,003* or *NC_000, 013.10:g.19,586,135_19,586,136insAB077488.1:129_366.*

2.5.5 Indels (Deletion/Insertions)

Indels, or deletion/insertions, of two or more consecutive nucleotides are described as a deletion followed by an insertion ("delins"): *g.27,106,792_27,106,805delinsGT* (NC_000,001.10), *c.6403_6409delinsGT* (NM_006,015.4), *r.6403_6409delinsgu* (NM_006,015.4), *p.Ile2135_-Leu2136delinsArg* (NM_006,015.4, NP_006,006.3).

2.5.6 Inversions

Inversions are described using "inv" after the position of the first and last nucleotides affected by the inversion: *g.27,106,792_27,106,805inv* (NC_000,001.10), *c.6403_6409inv* (NM_006,015.4), *r.6403_6409inv* (NM_006,015.4). Inversions do not occur on the protein level.

2.5.7 Conversions

Conversions are described using "con" after the position of the first and last residues affected by the conversion, followed by a description of the origin of the new nucleotides: *g.42,522,624_42,522,669con42536337_42,536,382* (NC_000,022.10), *c.1401_1446conNC_000,022.10:42,536,337_42,536,382* (NM_000,106.4), *r.1401_1446conNC_000, 022.10:42,536,337_42,536,382* (NM_000,106.4), *p.Asp11_His84conNP_005,201.2:Ala11_Pro84* (NM_014,617.3, NP_055,432.2).

2.5.8 Translocations

Although suggestions have been published to cover the description of translocations (Taschner and den Dunnen, 2011; Ordulu et al., 2014), official HGVS recommendations are not yet available. Historically, the description of translocations is the responsibility of the International Standing Committee on Human Cytogenetic Nomenclature (ISCN), and their recommendations are published in collaboration with the journal *Cytogenetic and Genome Research* (ISCN, 2013; Simons et al., 2013). Given the increased use of sequencing technologies to characterize chromosomal aberrations, the ISCN and HGVS committees are working on shared recommendations and a clear distinction between the two systems. The suggested separation is that when nucleotides are used in the description of a variant, the HGVS recommendations should be followed.

2.5.9 Variability in Repeated Sequences

Using the earlier discussed rules, the description of variants in stretches of repeated sequences can become rather complex. To simplify matters, such variants can be described using the position (range) of the first repeat unit and a number indicating the number of copies that are present: *g.146,993,570_146,993,572[14]* (NC_000,023.10), *c.-128_-126[14]* (NM_002,024.5). For short repeat units, an alternative is allowed using the positions and a description of the repeat unit: *g.146,993,570GGC[14]* (NC_000,023.10), *c.-128GGC[14]* (NM_002,024.5). Complex repeats can be described using the same principle: *g.146,993,570GGC[8]GGA[1]GGC[8]GGA[1]GGC[8]GGA[1]GGC[9]* (NC_000,023.10), *c.-128GGC[8]GGA[1]GGC[8]GGA[1]GGC[8]GGA[1]GGC[9]* (NM_002,024.5). When the size of the repeat unit is uncertain, this can be specified using parentheses: c.-128GGC[(600_800)] (NM_002,024.5).

2.5.10 Alleles

Humans are diploid organisms, i.e., most sequences are two copies (the autosomes, chromosomes 1-22). When more than one variant is described it should be made clear whether the variants were found on one chromosome (allele, DNA molecule) or on the two different chromosomes (alleles, maternal and paternal origin). Two or more changes compared to a reference sequence are described by combining the changes per molecule (allele) between square brackets ("[;];[;]") and using a semicolon (";") as the separator: *c.[76C>T;283G>C]* for two variants on one molecule (in cis); *c.[76C>T];[283G>C]* for two variants on two different molecules (in trans); and *c.[76C>T(;)283G>C]* for two variants where the phase is unknown.

2.5.11 RNA

As shown earlier, the description of variants at the RNA level follows that of DNA variants, using either a coding or noncoding RNA reference sequence (transcript). When a DNA variant yields several different transcripts on the RNA level, these variants are described between square brackets, separated by a "," (comma): *r.[76a>c, 70_77del].*

2.5.12 Protein

Variant descriptions at the protein level cover a few specific cases.

2.5.12.1 Nonsense Variants

Nonsense variants are described as substitutions, p.(Trp1298Ter) or p.(Trp1298*), not as deletions from the altered amino acid up to the translation stop codon (p.Trp1298_Leu3371del, Ter codon at 3318).

2.5.12.2 Frame Shifts

Variants that are predicted to shift the translational reading frame can be described using either a short or a long form: *p.(Arg97fs)* or *p.(Arg97ProfsTer23)*. In "fsTer#" (or "fs*#"), the symbol "#" indicates at which codon number in the new shifted reading frame a stop codon is encountered. The stop codon position in the new reading frame is calculated starting at the first amino acid that is changed by the frame shift, ending at the stop codon (Ter#). This stop codon position is never "1," since this should be described as a nonsense variant, like *p.(Trp1298Ter)* or *p.(Trp1298*)*.

2.5.12.3 Extensions

Variants in the translation start or stop codon, extending either the N- or C-terminal sequence, are described as extensions. Amino acids upstream of the original start site are numbered using a minus sign: *p.Met1ValextMet-12* (N-terminal extension of 12 amino acids, Met-12 to Thr-1). Similarly, *p.Ter110GlnextTer17* describes the observed extension of the C-terminus of the protein with 17 new amino acids as a consequence of a variant that changes Ter110 to Gln.

2.6 MUTALYZER

The stricter definition of the HGVS recommendations and its formalized description in Extended Backus-Naur Form (Laros et al., 2011) allowed the development of software tools that can check and/or generate HGVS descriptions (Fig. 2.1, Mutalyzer, Wildeman et al., 2008; Vis et al., 2015; HGVS Python package, Hart et al., 2014; Variant Validator, https://variantvalidator.org/). Note, however, that these tools are not yet without error, and their results should always be checked. In general, as can be expected, descriptions of simple changes are correct while those of more complex changes may contain errors.

A very powerful addition to the Mutalyzer suite is its "Variant Description Extractor" (VDE; Fig. 2.2, Vis et al., 2015). After copying or uploading the reference file or

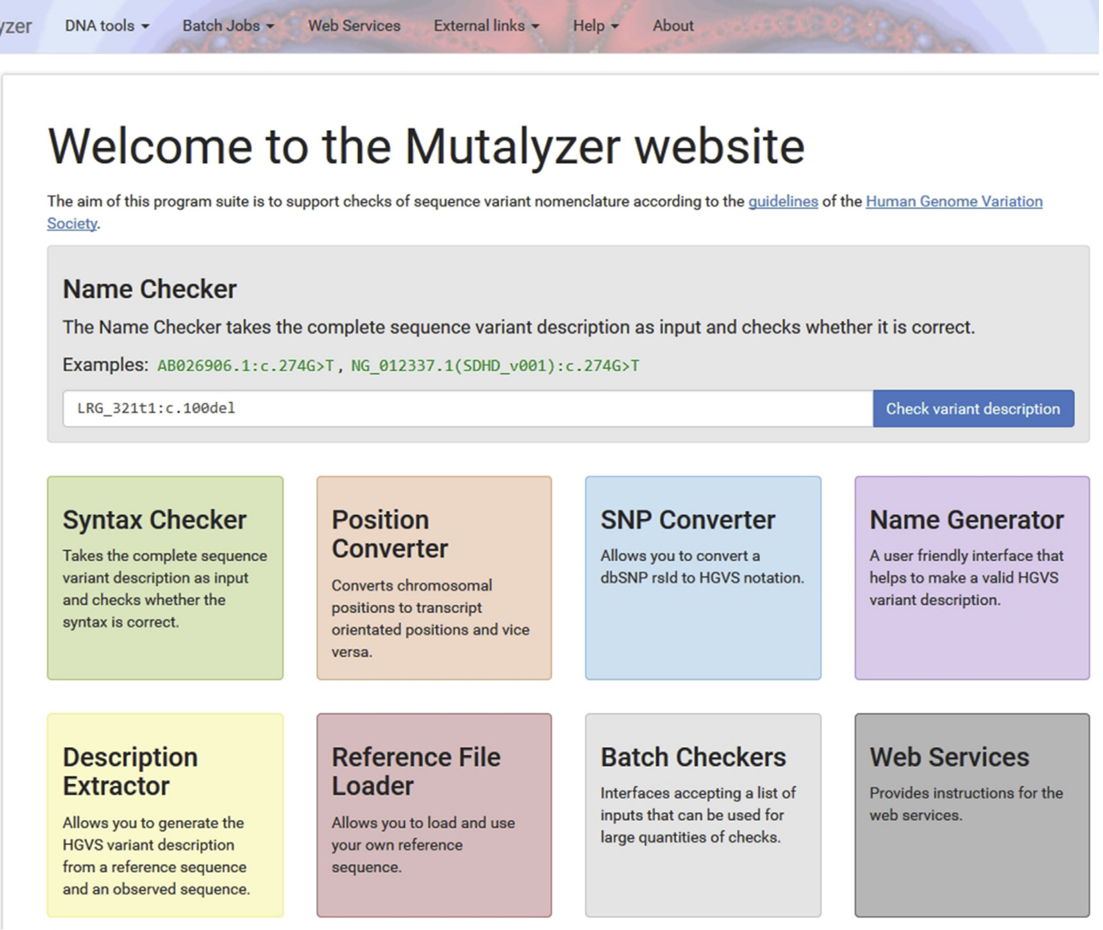


FIGURE 2.1 The Mutalyzer homepage provides access to the different functionalities.

(A)

Variant Description Extractor

Please note that this is an experimental service and we are currently limiting input sequences to 15,000 bp.

Extract the HGVS variant description from a reference sequence and an observed sequence. For now, we require the user to fill in two sequences. After the testing phase, we plan to use the underlying algorithm for:

- Disambiguation in the name checker. This will enable full support for complex variants.
- Comparison of two reference sequences. Useful for migrating a variant description to an other reference sequence.
- Implementation of a Reference Sequence Editor.

The algorithm is implemented in the HGVS variant description extractor. To apply it on longer input sequences than accepted on this page, you can download that package and run it locally.

Please supply a reference sequence and an observed sequence.

Reference input

- Enter a sequence (FASTA, FASTQ, or plain text).
- Upload a file (FASTA, FASTQ, or plain text).
- Enter a RefSeq accession number.

Reference sequence

ATGATGATCAGATACAGTGTGATACAGGTAGTTAGACAA

Example: ATGATGATCAGATACAGTGTGATACAGGTAGTTAGACAA

Sample input

- Enter a sequence (FASTA, FASTQ, or plain text).
- Upload a file (FASTA, FASTQ, or plain text).
- Enter a RefSeq accession number.

Sample sequence

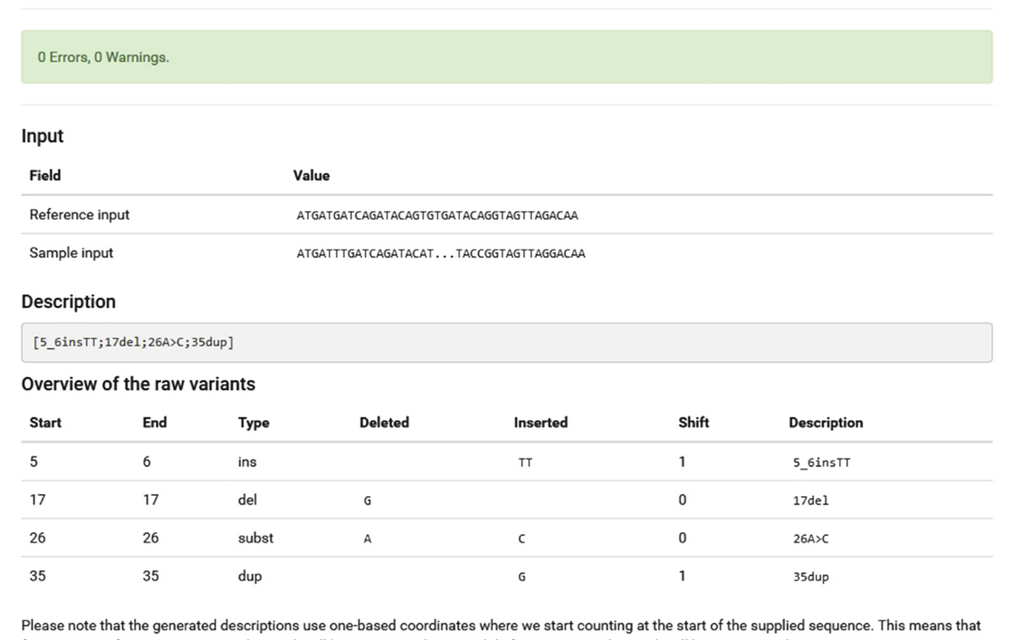
ATGATTTGATCAGATACATGTGATACCGGTAGTTAGGACAA

Example: ATGATTTGATCAGATACATGTGATACCGGTAGTTAGGACAA

Extract variant description

Help

(B)



Mutalyzer 2.0.17
released on 24 Mar 2016

HGVS nomenclature version 2.0

© 2009-2016 LUMC

FIGURE 2.2 Mutalyzer's Variant Description Extractor (Vis et al., 2015) compares the specified reference and observed sequences (A) to generate the HGVS variant description (B).

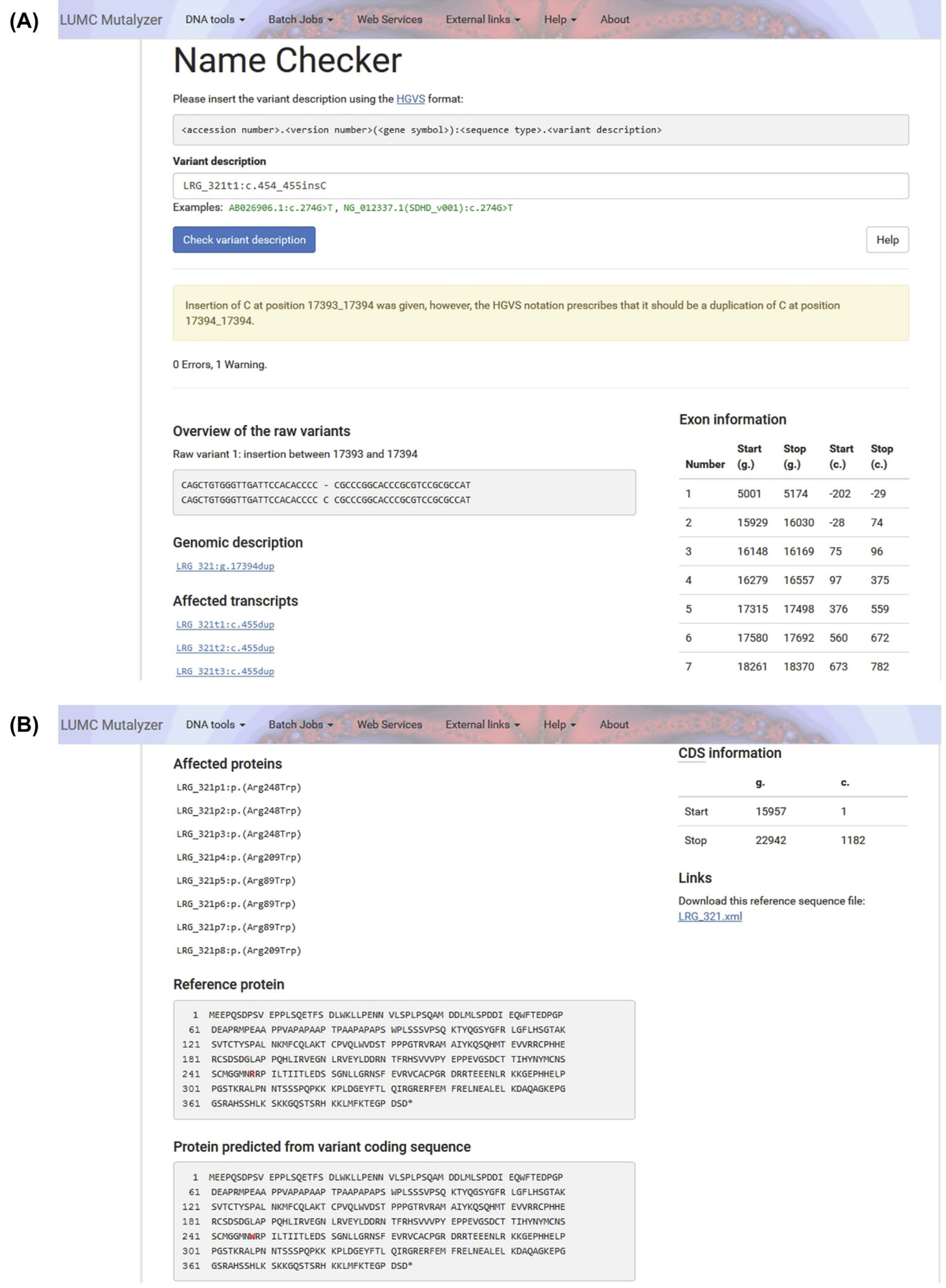


FIGURE 2.3 Mutalyzer's Name Checker verifying and correcting the description of the c.454_455insC variant in transcript t1 of the *TP53* reference sequence LRG_321. The Name Checker returns the correct description, c.455dupC on transcript t1, and the correct descriptions on the other annotated transcripts (A) as well as the predicted consequences on their corresponding proteins (B).

entering a RefSeq accession number in one window and the variant sequence in the second window, the VDE compares both sequences and generates the variant description(s). Other Mutalyzer tools include the following:

- Name Generator: a user interface that helps to make a valid HGVS variant description
- Name Checker: checks whether the sequence variant description is correct (Fig. 2.3)
- Position Converter: converts chromosomal positions to transcript-based positions and vice versa (Fig. 2.4)
- Syntax checker: checks whether the syntax of the sequence variant description is correct
- SNP Converter: converts a dbSNP rsID to HGVS notation (genomic, coding DNA, and protein)

All tools are available to analyze variants one by one or in batch mode. Mutalyzer has a web service (SOAP and HTTP/RPC+JSON) and is Open Source, available under the GNU Affero General Public License.

2.7 CONCLUDING REMARKS

Over the years, the sequence variant nomenclature system proposed by the HGVS has developed into an internationally accepted standard. The HGVS recommendations facilitate consistent and unambiguous description of sequence variants, an essential requirement in reporting and exchanging information on the analysis of a genome. The recommendations are currently commissioned through the SVD-WG operating under the auspices of the HGVS, HVP,

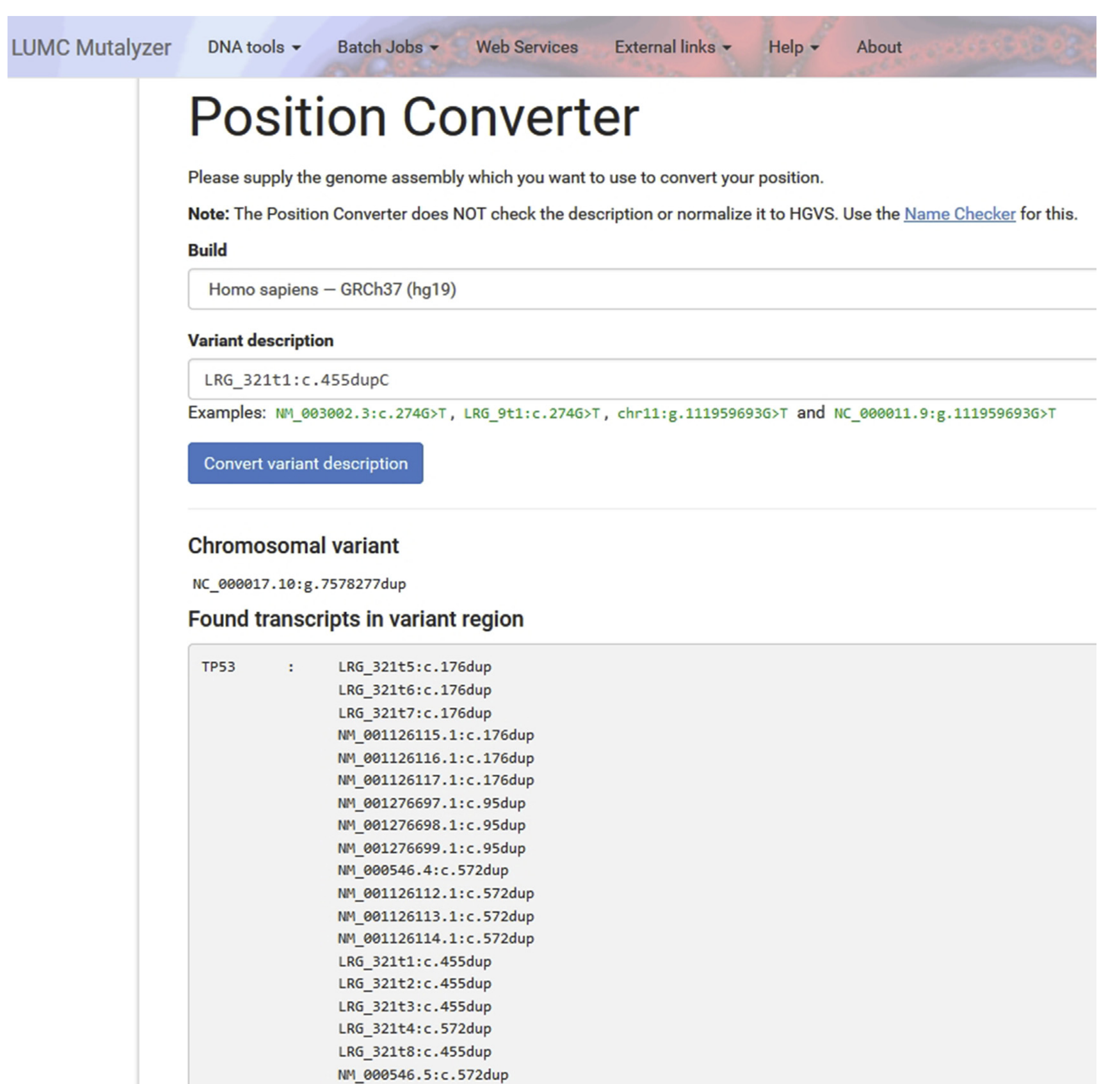


FIGURE 2.4 Mutalyzer's Position Converter supports the conversion of chromosomal positions to coding DNA positions and vice versa. Given a single coding DNA variant description, it will return the corresponding chromosomal position for the selected genome build and variant positions for all annotated overlapping Locus Reference Genomic (LRG) and transcript reference sequences. It does not check the variants, so for completeness it should be combined with the Name Checker.

and HUGO. Efforts of the SVD-WG have focused on removing inconsistencies and tightening definitions, allowing for automatic data processing (den Dunnen et al., 2016). While the recommendations do not yet cover all possible variant types, e.g., a proposal covering translocations is currently under discussion, they are mature and near complete. Additional requests are still received, e.g., to cover exon numbering, RNA editing, DNA modification (methylation), protein modification (phosphorylation, glycosylation), etc., raising the question as to how far the recommendations should extend. In this chapter we have presented the basics of the HGVS recommendations but not every tiny detail. For a more extensive overview we refer to the online version of the recommendations and the material available there (http://www.HGVS.org/varnomen/).

ACKNOWLEDGMENTS

We gratefully acknowledge everyone who has, over the years, contacted the Human Genome Variation Society to point out problems with the recommendations, errors/inconsistencies on the webpages, and those who have participated in the recent round of community consultation of the Sequence Variant Description Working Group (SVD-WG). In addition, we thank the Human Variome Project (International Coordinating Office) for administrative support establishing and operating the SVD-WG.

REFERENCES

Ad Hoc Committee on Mutation Nomenclature, 1996. Update on nomenclature for human gene mutations. Hum. Mutat. 8, 197–202.

Antonarakis, S.E., 1998. Recommendations for a nomenclature system for human gene mutations. Hum. Mutat. 11, 1–3.

Beaudet, A.L., Tsui, L.C., 1993. A suggested nomenclature for designating mutations. Hum. Mutat. 2, 245–248.

Dalgleish, R., Flicek, P., Cunningham, F., Astashyn, A., Tully, R.E., Proctor, G., Chen, Y., McLaren, W.M., Larsson, P., Vaughan, B.W., Béroud, C., Dobson, G., Lehväslaiho, H., Taschner, P.E.M., den Dunnen, J.T., Devereau, A., Birney, E., Brookes, A.J., Maglott, D.R., 2010. Locus Reference Genomic sequences: an improved basis for describing human DNA variants. Genome Med. 2, 24.1–24.7.

den Dunnen, J.T., Antonarakis, S.E., 2000. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum. Mutat. 15, 7–12.

den Dunnen, J.T., Dalgleish, R., Maglott, D.R., Hart, R.K., Greenblatt, M.S., McGowan-Jordan, J., Roux, A.F., Smith, T., Antonarakis, S.E., Taschner, P.E.M., 2016. Update: the HGVS recommendations for the description of sequence variants. Hum. Mutat. 37, 564–569.

Gulley, M.L., Braziel, R.M., Halling, K.C., Hsi, E.D., Kant, J.A., Nikiforova, M.N., Nowak, J.A., Ogino, S., Oliveira, A., Polesky, H.F., Silverman, L., Tubbs, R.R., Van Deerlin, V.M., Vance, G.H., Versalovic, J., Molecular Pathology Resource Committee, College of American Pathologists, 2007. Clinical laboratory reports in molecular pathology. Arch. Pathol. Lab. Med. 131, 852–863.

Hart, R.K., Rico, R., Hare, E., Garcia, J., Westbrook, J., Fusaro, V.A., 2014. A Python package for parsing, validating, mapping and formatting sequence variants using HGVS nomenclature. Bioinformatics 31, 268–270.

ISCN, 2013. Shaffer, L.G., Jean McGowan-Jordan, J., Michael Schmid, M. (Eds.) 2013. Karger Publishers.

IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN), 1984. Nomenclature and symbolism for amino acids and peptides. Recommendations 1983. Eur. J. Biochem. 138, 9–37.

Laros, J.F., Blavier, A., den Dunnen, J.T., Taschner, P.E.M., 2011. A formalized description of the standard human variant nomenclature in Extended Backus-Naur Form. BMC Bioinforma. 12, S4–S5, 1–7.

MacArthur, J.A.L., Morales, J., Tully, R.E., Astashyn, A., Gil, L., Bruford, E.A., Larsson, P., Flicek, P., Dalgleish, R., Maglott, D.R., Cunningham, F., 2014. Locus Reference genomic: reference sequences for the reporting of clinically relevant sequence variants. Nucleic Acids Res. 42, D873–D878.

Ordulu, Z., Wong, K.E., Currall, B.B., Ivanov, A.R., Pereira, S., Althari, S., Gusella, J.F., Talkowski, M.E., Morton, C.C., 2014. Describing sequencing results of structural chromosome rearrangements with a suggested next-generation cytogenetic nomenclature. Am. J. Hum. Genet. 94, 695–709.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H.L., ACMG Laboratory Quality Assurance Committee, 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. 17, 405–424.

Simons, A., Shaffer, L.G., Hastings, R.J., 2013. Cytogenetic nomenclature: changes in the ISCN 2013 compared to the 2009 edition. Cytogenet. Genome Res. 141, 1–6.

Taschner, P.E.M., den Dunnen, J.T., 2011. Describing structural changes by extending HGVS sequence variation nomenclature. Hum. Mutat. 32, 507–511.

Vis, J.K., Vermaat, M., Taschner, P.E., Kok, J.N., Laros, J.F., 2015. An efficient algorithm for the extraction of HGVS variant descriptions from sequences. Bioinformatics 31, 3751–3757.

Wildeman, M., van Ophuizen, E., den Dunnen, J.T., Taschner, P.E.M., 2008. Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker. Hum. Mutat. 29, 6–13.

Chapter 3

Low- and Medium-Throughput Variant Detection Methods: A Historical Perspective

A. Sgourou[1], A. Papachatzopoulou[2], T. Katsila[2] and G.P. Patrinos[2]
[1]Hellenic Open University, Patras, Greece; [2]University of Patras School of Health Sciences, Patras, Greece

3.1 INTRODUCTION

Molecular diagnostics primarily aim to detect genomic variants, also known as single nucleotide polymorphisms (SNPs), which can be either pathogenic or benign. Since the advents of the polymerase chain reaction (PCR) (Mullis and Faloona, 1987) and the thermostable *Taq* polymerase (1998; see also Chapter 1), a number of PCR-based genetic testing methods have been described. Those approaches were all low-throughput, for example, allowing the detection of a single or just a few genomic variants. These methods can be summarized in two main categories: genetic screening methods, specifically designed to detect specific genomic variant(s), and genetic scanning methods, aiming to detect every genomic variant within the PCR-amplified fragment under study. This chapter aims to provide an overview of the most commonly used low-throughput genetic testing methodologies, which served as the basis for the development of the medium- and high-throughput approaches used today in modern molecular diagnostic laboratories.

3.2 GENETIC SCREENING METHODS

As described in Chapter 1, the very first approach that was used for diagnosing a genetic disease at the molecular level was PCR coupled with restriction fragment length polymorphism (RFLP) analysis to detect the variant leading to sickle cell disease. Since then, there have been several other approaches that are dependent upon allele-specific PCR-based DNA amplification to screen for known genomic variants. Allele-specific detection methods have been widely used in research and molecular diagnostics since their initial development in the late 1980s, not only because they are easily applicable to the analysis of virtually any known variant but also because these methods do not require expensive and sophisticated instrumentation. These approaches, which have constituted the foundation for several research and diagnostic laboratories in the past, are the amplification refractory mutation system (ARMS), allele-specific oligonucleotide (ASO) probes, and oligonucleotide ligation assays (OLAs). There have also been other genetic screening approaches, such as competitive oligopriming, chemical and enzymatic cleavage, etc., that have been less frequently employed for molecular diagnostics and as such will not be described herein.

3.2.1 Amplification Refractory Mutation System

ARMS is based on the principle that a mismatch between the 3′ nucleotide of a PCR primer and the template prevents primer extension by the *Taq* polymerase. In 1989, several independent groups described this approach for analyzing known point mutations in DNA and distinguishing between normal, heterozygous, and homozygous mutant genotypes (Newton et al., 1989; Nichols et al., 1989; Okayama et al., 1989; Sommer et al., 1989; Wu et al., 1989). Amplification of the normal allele, and not that of the mutant, is accomplished using a primer that is complementary to the normal allele and has a mismatch between the 3′ residue and the mutant allele. Conversely, only the mutant will be amplified if the 3′ residue of the

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00003-1
Copyright © 2017 Elsevier Ltd. All rights reserved.

primer is complementary to the mutant allele and not the normal allele. The basic concept and an indicative application are illustrated in Fig. 3.1A and B, respectively.

The design and optimization of the PCR−ARMS protocols rely on the target sequence and the nucleotide differences that define the alleles. In addition to mismatches between the 3′ terminal base of the primer and the target, single mismatches should be incorporated at several positions from the 3′ terminus. The design and optimization of PCR−ARMS primers follow the same considerations used for any other type of PCR, namely: (1) primers are chosen to have comparable theoretical melting temperatures (T_m), (2) primer lengths are generally 20 nucleotides or longer, although the length is less important than the T_m, and (3) primers should not have self-complementary sequences of four nucleotides or more nor should they have more than four-nucleotide complementarity between their 3′ ends. As with any conventional PCR strategy, avoiding false

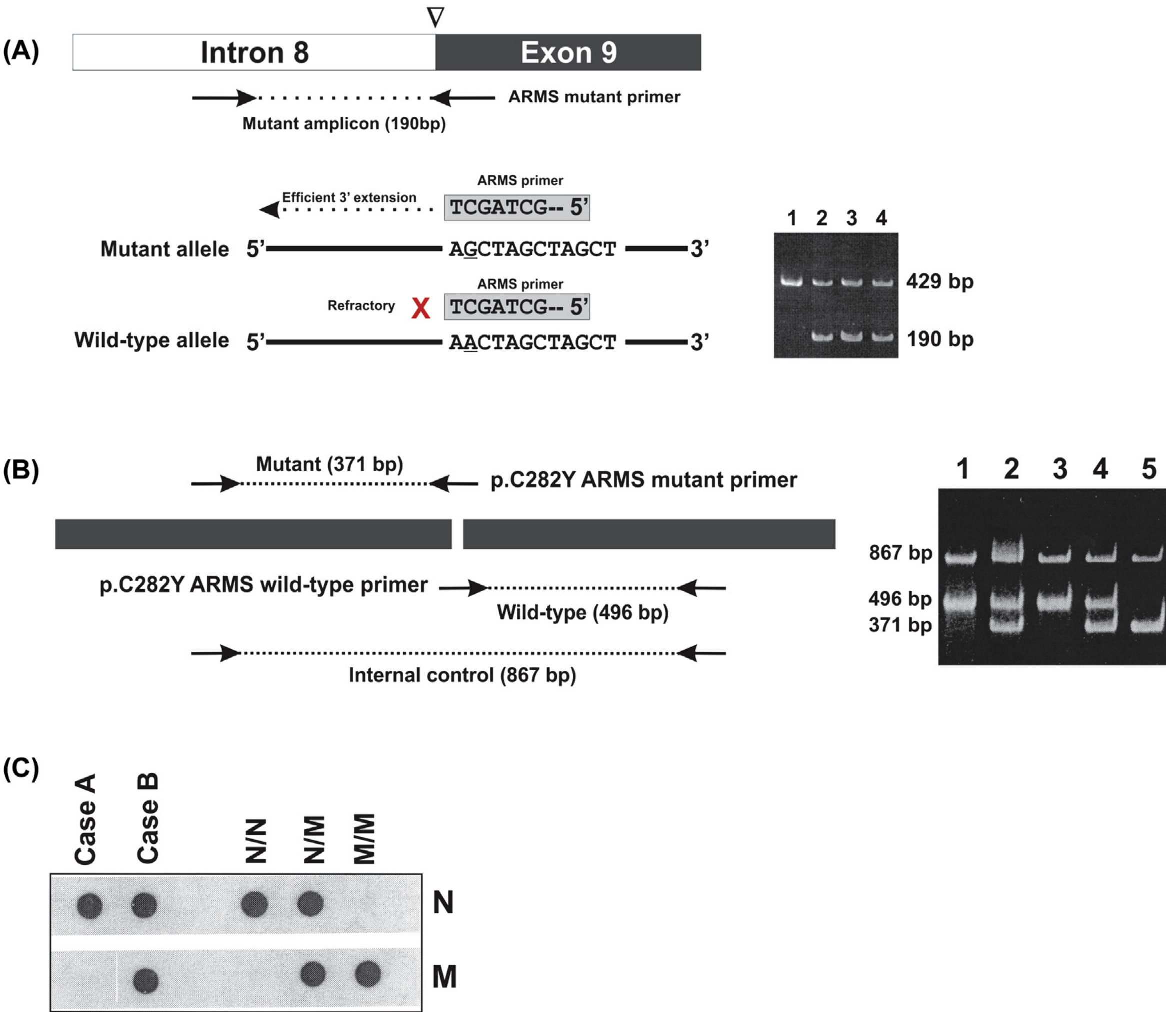


FIGURE 3.1 (A) PCR−ARMS detection of a single *DHCR7* mutation. Schematic representation of the boundary between intervening sequence 8 and exon 9 of the *DHCR7* gene showing the position of a common mutation causing Smith-Lemli-Opitz syndrome. The mutation alters the canonical splice acceptor sequence (AG > AA) and can be detected using an ARMS primer specific for the mutant allele. The analysis of PCR−ARMS products by nondenaturing polyacrylamide gel electrophoresis, visualized by ethidium bromide staining and UV fluorescence, reveals a 190-bp fragment specific for the IVS 8-1 G > C variant allele. The 429-bp fragment corresponds to a region of the *HFE* gene that serves as a positive internal control for PCR amplification. Lane 1: normal control; lanes 2−4: IVS 8-1 G > C heterozygote. (B) Genotyping of *HFE* p.C282Y mutation using PCR−ARMS. Schematic representation of a portion of the *HFE* gene showing the most common mutation associated with hereditary hemochromatosis (p.C282Y). The multiplex analysis contains primers for two different PCR−ARMS assays, run off opposite strands. One PCR−ARMS assay is specific for the mutant allele, whereas the other assay is specific for the normal allele. The PCR products for the mutant and normal alleles have different lengths and can easily be distinguished by gel electrophoresis. Lanes 1, 3: normal samples; lanes 2, 4: p.C282Y heterozygotes; lane 5: p.C282Y homozygote. (C) Forward PCR−ASO of *HBB* gene mutations. PCR−ASO analysis of the most common *HBB* variant in the Hellenic population (IVS I-110 G > A), leading to β-thalassemia. N/N: homozygous normal sample; N/M: heterozygous mutant sample; M/M: homozygous mutant sample; Case A: homozygous normal sample; Case B: heterozygous mutant sample.

negative results can be overcome by incorporating a degenerate nucleotide into the primer or by targeting the opposite strand for amplification.

If several sequence variants need to be detected in a multiplex ARMS detection protocol, one first needs to optimize the PCR conditions for the sensitive and specific detection of each allele, namely adjusting the PCR cycling parameters and $MgCl_2$ concentration. Subsequently, and once those parameters are set for each allele, the primer pairs can be combined to evaluate the performance of the multiplex ARMS assay. Following DNA amplification, agarose or polyacrylamide gel electrophoresis systems can be used to resolve the wild type from the variant amplicon, which should differ in sizes to be clearly distinguished during electrophoresis. The specificity of PCR−ARMS is such that a large number of samples can be analyzed if needed in pools, as the approach is capable of detecting a single positive sample in pools of 30 or more samples (Nowaczyk et al., 2001; Waye et al., 2002).

PCR−ARMS assays can be developed, validated, implemented, and multiplexed rather easily, which constitutes the main advantage of this approach, while they do not require expensive and sophisticated detection systems (Liu et al., 2016; Medicare et al., 2016). On the contrary, as with other variant screening methods, PCR−ARMS can only be used to detect known genomic variants.

3.2.2 Allele-Specific Oligonucleotide Hybridization

The analysis of single nucleotide variations in DNA using hybridization with ASO probes is based on the principle that even single nucleotide mismatches between a probe and its target can destabilize the hybrid. ASO probes can be designed to be complementary and specific for the various alleles, thus providing a simple methodology to detect any known mutation or SNP.

The use of ASO probes actually precedes PCR and was a commonly used approach to analyze cloned DNA, using radioactively labeled ASO probes to hybridize on genomic DNA that has been immobilized on a membrane after restriction endonuclease digestion and electrophoresis (Conner et al., 1983; Orkin et al., 1983; Pirastu et al., 1983). With the advent of PCR amplification (Saiki et al., 1985, 1988a,b), PCR−ASO became one the methods of choice to analyze known point mutations within amplified DNA fragments (Saiki et al., 1986).

ASO probes are generally short oligonucleotides (15−17 mers) with a low GC content (or guanine-cytosine content) (usually 30−50%), designed such that the discriminating nucleotide is located somewhere in the middle of the probe. The right choice of the ASO probes requires a significant time investment and testing of several candidate probes that must be evaluated individually using both positive and negative control samples, all under fixed hybridization and wash stringency conditions. Also, caution should be exercised to avoid sequences that are polymorphic, which may lead to false negative results if a polymorphism impairs annealing and/or destabilizes the ASO/target hybrid.

There are two different variations of the theme of PCR−ASO: the forward ASO format and the reverse ASO format.

In the forward ASO format, the PCR-amplified DNA fragments are immobilized onto a filter or membrane, and labeled oligonucleotide probes that are complementary and specific for a given DNA sequence are hybridized to the filter. Subsequently, the filter is washed at the appropriate stringency to dissociate any probe molecules and exposed. The first PCR−ASO protocols used radiolabeled oligonucleotide probes with [γ-^{32}P] and X-ray film exposure to detect the membrane-bound probe-target hybrids (Saiki et al., 1986). Alternative protocols used biotinylated ASO probes and probe-target hybrid detection using streptavidin conjugated with horseradish peroxidase (Saiki et al., 1988b), which can be detected either using a colorimetric detection with tetramethylbenzidine and hydrogen peroxide or chemiluminescent substrates. The forward ASO format is particularly useful when large numbers of samples need to be analyzed for a small number of variant alleles (Fig. 3.1C). For a larger number of alleles being tested, the forward ASO format gets more cumbersome, requiring separate labeled probes and hybridization cycles for each allele being tested.

For this reason, the reverse ASO format was conceived to circumvent this problem. In reverse ASO, different ASO probes get immobilized onto a filter and/or membrane and hybridization of the filter to labeled PCR-amplified DNA occurs. The original reverse ASO format (also known as reverse dot blot) employed probes that had poly(dT) tails added to their 3′ termini and were immobilized on the nylon membranes by UV cross-linking (Saiki et al., 1989). The method was subsequently improved by covalent binding of the ASO probes to membranes via 5′ amino linkers (Zhang et al., 1991; Chehab and Wall, 1992). The reverse ASO format has become a widely used tool for the routine screening of genes that have numerous mutant alleles (reviewed in Gold, 2003), and several reverse dot-blot test kits have been made available for mutation screening for several genetic diseases, including α-thalassemia (Chan et al., 1999; Foglietta et al., 2003), β-thalassemia (Chehab, 1993; Cai et al., 1994), and cystic fibrosis (Chehab and Wall, 1992; Makowski et al., 2003). Also, several commercial kits have been developed for α- and β-thalassemia mutation detection (reviewed in Patrinos et al., 2005).

The PCR−ASO method, particularly the reverse format, provides a convenient approach for simultaneously screening large numbers of genomic variants, can be

applied to any known sequence variation, and does not require any sophisticated instrumentation for variant detection. On the contrary, the amount of work required for standardizing the hybridization and wash stringency condition for the simultaneous hybridization of several oligonucleotide probes is a significant drawback, particularly for small laboratories with limited resources, requiring a significant initial investment.

3.2.3 Oligonucleotide Ligation Assays

The principle of OLA lies in the ability of DNA ligases to join short oligonucleotides covalently to each other, which indirectly reflects the genotype of the target DNA. The joining of two oligonucleotide probes by DNA ligase is dependent on three different conditions: (1) hybridization of the probes to the target DNA, (2) the probes must lie directly adjacent to each other in a 5′ to 3′ orientation, and (3) the probes must have perfect base-pair complementarity with the target DNA at the site of their joining. If all three conditions are fulfilled, then sequence variants can be reliably, accurately, and reproducibly discriminated using OLA. The reaction can be monitored by introducing a detectable function in either of the probes in the ligation reaction. Several different analytical techniques are in use, differing from each other in their detection methods and instrumentation needs.

In 1988, this enzymatic reaction was applied to ligate short single-stranded DNA probes as a means to detect sequence variants (Alves and Carr, 1988; Landegren et al., 1988). This was the beginning of the development of several analytical methods utilizing DNA ligases, for example, OLA and ligase chain reaction. When thermostable DNA ligases became commercially available in 1990, they enabled the temperature cycling of the ligation reaction. OLA was revolutionized in 1990 when PCR was coupled to the assay prior to the ligation reaction (Nickerson et al., 1990), and nowadays the method is used in numerous clinical and research laboratories worldwide for the detection of a wide variety of medically relevant mutations.

In most applications, the relevant gene fragment is first amplified in a PCR, although in some applications genomic DNA has been directly subjected to the ligation reaction. The template DNA for OLA is denatured to form a single-stranded target DNA. Two adjacent oligonucleotide probes (approximately 20 mers) and a DNA ligase enzyme are needed for each ligation reaction. The oligonucleotides for OLA are designed to juxtapose at a previously identified mutation site on the template. The probes are hybridized to the target DNA in a head-to-tail (5′ to 3′) orientation. If the probes and the target DNA are perfectly complementary to each other, the DNA ligase enzyme forms a phosphodiester bond between the probes. If there is even a single base pair (bp) mismatch on either side or close to the nick between the probes, the efficiency of the enzyme is significantly decreased, and the ligation is prevented; therefore it can be distinguished from a perfect match (Wu and Wallace, 1989; Pritchard and Southern, 1997). For biallelic sequence variants, three synthetic oligonucleotide probes are used: one allele-specific probe is needed for each allele to be detected and a third, common probe then hybridizes to the target DNA sequence immediately adjacent to the allele-specific probes and is completely complementary to either allelic target. If a thermostable DNA ligase is used in the reaction, the denaturation and the ligation steps can be cycled to achieve a linear amplification of the ligated probes, thus enhancing the sensitivity of the assay.

There are several detection methods, and the choice between them is dependent on many factors, such as the available instrumentation in the laboratory, the number of samples to be analyzed, and the number of genetic variations to be analyzed per sample, as well as costs and the need for speed, simplicity, and robustness. There are assays based on the solid-phase capture of ligated probes, where one of the oligonucleotide probes used for the ligation reaction can be attached directly to a solid-phase support, for example, paramagnetic particle or biosensor chip (Martinelli et al., 1996; Zhong et al., 2003). However, a more common approach is to capture the probe on solid-phase by hybridization or streptavidin−biotin interaction (Nickerson et al., 1990; Iannone et al., 2000). There are several detection methods utilizing 96- or 386-well microplates. Detection on microplates can be done colorimetrically, using a colorimetric enzyme-linked immunosorbent assay-based detection (Nickerson et al., 1990), chemiluminometrically (Tannous et al., 2003), or using time-resolved fluorometry (Hansen et al., 1995). There are other less frequently used detection methods, such as fluorometric detection on microspheres (Iannone et al., 2000) and colorimetric detection nonbiosensor arrays, raising the assay's throughput to an even higher level (Zhong et al., 2003). There are also electrophoretic assays, based on the fact that the probe size is raised upon ligation. As such, the ligation products are resolved from each other and from the nonligated probes under denaturing conditions on high-resolution polyacrylamide gels or capillary electrophoresis by their unique electrophoretic mobility. The detection method can be either autoradiography or fluorescence, depending on the label used in the common probes (Day et al., 1995a; Grossman et al., 1994).

Several applications of OLA for the detection of molecular defects causing human diseases are, even today, in use in molecular diagnostic laboratories on a global scale. The first report describing OLA as a molecular diagnostic method is for the diagnosis of sickle cell disease (MIM 603903; Landegren et al., 1988). Since then OLA has been used for prenatal screening of this disease (Day et al.,

2002). Also, OLA has been used to detect a wide variety of medically relevant mutations causing inherited diseases or mutations contributing to a common disease, such as cystic fibrosis, hereditary hemochromatosis, and variants occurring in isolated populations, such as the Finnish Disease Heritage. Although most of the OLA applications are developed and validated in-house, several commercial genotyping kits are available for the diagnosis of inherited disorders, such as kits for detecting the common genetic variants of clotting factors II (prothrombin) and V, as well as methylene tetrahydrofolate reductase, which predisposes an individual to thrombophilia. Even these commercial kits have to be validated for clinical use in each laboratory.

One of the main advantages of PCR−OLA is that it reliably distinguishes heterozygous and homozygous mutant cases from homozygous normal individuals. Also, the use of thermostable DNA ligase improves the specificity of the reaction since the ligation can be performed at higher temperatures, thus increasing the annealing stringency (Barany, 1991). Since the assay can be performed over a broad range of ligation temperatures, salt concentrations, and DNA ligase concentrations without any loss of sensitivity (Hansen et al., 1995), this eliminates the risk of false-negative or false-positive reactions due to variations in these factors. Also, there are several means for the detection of ligation products, such as colorimetry, chemiluminometry, and time-resolved fluorometry, that can be performed in the microplate format by immobilizing one of the probes directly or by biotin−streptavidin interaction into a well, a particle, or a pin. Furthermore, PCR−OLA can be multiplexed for the detection of several alleles or loci in the same reaction. However, there are also some limitations, such as a lack of reproducibility, owing to the generation of spurious amplification products in the PCR step (Fox et al., 2007).

3.3 GENETIC SCANNING METHODS

Contrary to the genetic screening methods described earlier that are able to detect only known variants for which they are designed, one should ensure that the PCR-amplified fragment under study is thoroughly investigated for the existence of any possible variant, both known as well as unknown. As such, other genotyping methods, also known as genetic scanning methods, should be used that are able to detect the existence of any possible variant in the fragment in question. The most popular genetic scanning approaches that constituted the cornerstone of any molecular diagnostics laboratory in the last decade were single-strand conformation polymorphism (SSCP), heteroduplex analysis (HDA), and denaturing- and temperature-gradient gel electrophoresis (DGGE and TGGE, respectively). In the following paragraphs, we will summarize the earlier mentioned methods, keeping in mind that those techniques were designed to scan any PCR-amplified fragment for single point variants and small indels. There are other methods that are designed to screen for larger deletions, insertions, and gross rearrangements, such as multiplex ligation-dependent probe amplification, which lie outside the scope of the present chapter.

3.3.1 Single-Stranded Conformation Polymorphism Analysis

SSCP is perhaps the most popular electrophoresis-based genetic scanning method. Coupled to DNA amplification of the sequence to be analyzed, these techniques have become the methods of choice for several molecular diagnostic laboratories. This can be explained mainly by the numerous advantages, namely their technical simplicity and relatively high specificity for the detection of sequence variations, the low operation costs, and the potential for automation for high-throughput mutation analysis. If fluorescently labeled primers are employed during DNA amplification, SSCP analysis can be performed in gel- or capillary electrophoresis-based automated sequencers, hence allowing for precise, reproducible, and high-throughput analysis of the genomic variation.

The analysis of SSCP has been established by Orita et al. (1989) as a simple, efficient, and reliable method for the detection of sequence alterations in genomic loci. Based on PCR, SSCP was developed soon after the introduction of PCR technology and relied on the fact that relatively short single-stranded DNA fragments can migrate in a nondenaturing gel not only as a function of their size but also of their sequence. In other words, following the amplification of any given DNA sequence, the amplified DNA fragments are subjected to denaturation with either heat or chemical agents, such as formamide. Subsequently, the denatured DNA fragments are electrophoresed through a native (nondenaturing) polyacrylamide gel. During electrophoresis, single-stranded DNA fragments adopt a specific three-dimensional shape according to their nucleotide sequence and hence exhibit a unique conformation. Therefore their electrophoretic mobility is dependent upon the previously mentioned three-dimensional shape. Based on these principles, it is well understood that even a single base difference between a DNA fragment being tested and its wild type counterpart is sufficient to adopt a different conformation and thus to migrate at a different position during electrophoresis (Fig. 3.2A).

Fluorescent SSCP (F-SSCP) is a variation of the theme of the standard SSCP method. F-SSCP is a nonradioactive, high-resolution PCR−SSCP method in which fluorescently labeled PCR products are electrophoresed and detected by an automated DNA sequencer (Makino et al., 1992). Coupling SSCP analysis with an automated DNA

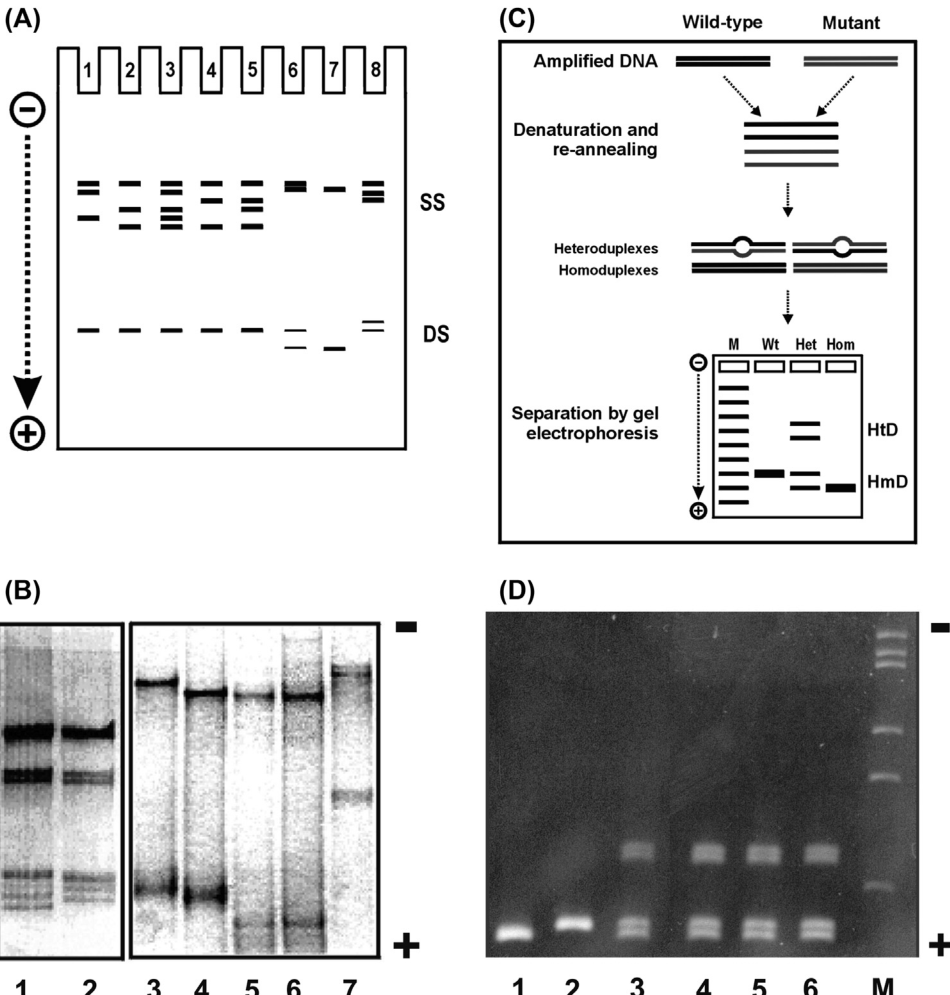


FIGURE 3.2 (A) Schematic drawing of a typical result from a single-strand conformation polymorphism (SSCP) analysis. After staining, the gel can be divided into two parts: the part with the bands corresponding to the single-stranded (SS) alleles and the part with the double-stranded (DS) bands, as complete denaturation is sometimes not feasible. In this example, lane 1 corresponds to the wild type pattern for any given locus analyzed. If lane 2 corresponds to a homozygous case for mutation A, then lane 3 corresponds to a heterozygous case for the same mutation, and lanes 4 and 5 correspond to a homozygous case for mutation B and a compound heterozygous case for mutations A and B, respectively. Lanes 6−8 are characteristic examples of the electrophoretic pattern of small deletions or insertions, which can also be distinguished at the DS part of the gel (see also Fig. 5.5b). Lanes 6, 7: heterozygous and homozygous cases for a small deletion, respectively; Lane 8: heterozygous case for a small insertion. (B) Analysis of the coding region of the human *G6PD* locus (*Adapted from Menounos, P., Zervas, C., Garinis, G., Doukas, C., Kolokithopoulos, D., Tegos, C., Patrinos, G.P., 2000. Molecular heterogeneity of the glucose-6-phosphate dehydrogenase deficiency in the Hellenic population. Hum. Hered. 50, 237−241.*). Lanes 1, 2: Exon 8 (lane 1: normal individual; lane 2: heterozygous case); Lanes 3−7: Exon 11 (lanes 5, 6: wild type pattern; lanes 4, 7: heterozygous cases for different mutations; lane 3: heterozygous case with two point mutations *in cis*, one of which is the same as in lane 4 [note the minor mobility difference of the bands for the SS alleles between lanes 3 and 4]). (C) The heteroduplex analysis (HDA) principle. Unlike SSCP analysis, amplified DNA is now subjected to denaturation followed by a slow reannealing of the denatured alleles, leading to both homoduplexes (HmD) and heteroduplexes (HtD). The latter migrate slower during gel electrophoresis, due to their sequence mismatch(es), and therefore heterozygous (Het) and homozygous (Hom) cases can be easily distinguished from the wild type (Wt) based on their electrophoretic pattern. M: size marker. (D) Mutation screening in the exon 10 of the *CFTR* gene for the p.F508del mutation, leading to cystic fibrosis, using HDA. Electrophoresis is performed in a nondenaturing 8% acrylamide gel. The amplification product of the wild type allele is 97 bp and that of the mutant allele is 94 bp. Lanes 3−6 correspond to p.F508del heterozygotes, lane 1 corresponds to a p.F508del homozygote, and lane 2 corresponds to a wild type individual. The electrophoretic mobility of the heteroduplexes is retarded compared to the homoduplexes. M: φX174/HaeIII size marker. *Photo courtesy of Dr. Angeliki Balasopoulou, Athens, Greece.*

sequencer enables highly sensitive fragment detection and permits the strict control of any desired temperature, a fundamental requirement of SSCP analysis to obtain reproducible results. Coelectrophoresis of an internal size standard DNA labeled with a dye different from that of the sample DNA allows the relative fragment mobility of sample DNAs to be reproducibly determined using the mobility of the internal standard DNA as a reference.

There are several parameters that influence SSCP analysis that will eventually affect electrophoretic migration:

1. *DNA amplification* (such as the size of PCR products) with a PCR fragment size of approximately 550 bp (Hayashi, 1991; Hayashi and Yandell, 1993).
2. *Denaturation*: In SSCP analysis, it is important to achieve complete and as much irreversible denaturation of the DNA strands as possible. Usually, denaturation of the PCR products is done in 95°C for 5−7 min, followed by an incubation step at 4°C for approximately 10 min. Alternatively, there are numerous denaturing agents, such as formamide, methyl-mercuric hydroxide, sodium hydroxide, and urea, which seem to perform well (Humphries et al., 1997).
3. *Electrophoretic conditions*, namely the length and duration of gel run, the ionic strength in the buffer being used, the temperature, and the gel matrix composition. It must be noted that there is neither adequate theory nor any physicochemical model available that could allow one to predict the three-dimensional structure of any given single-stranded DNA fragment and, as a result, its electrophoretic mobility. Apart from the size of the DNA fragment and its GC content, the following parameters have been empirically found to affect the sensitivity of SSCP analysis: the gel matrix composition, the buffer composition (ionic strength, the pH and buffer supplements, such as glycerol), the duration of the gel run, the gel length, the DNA concentration, and the electrophoresis temperature.
4. *Detection*: Visualization of the single-stranded DNA fragments can be done using a number of detection methods. This includes the use of radioactivity and some of the other described protocols that utilize silver staining or detection with fluorescent dyes. Alternatively, in F-SSCP, analysis can be done on automated sequencers with fluorescent dye-labeled DNA fragments. From those, the silver staining approach comprises one of the most straightforward, fast, and sensitive methods for the visualization of bands on conventional SSCP gels.

Several factors influence sensitivity in SSCP analysis and therefore need to be taken into account in order to obtain reproducible results as well as to maximize the sensitivity of the mutation detection. Optimization of these factors is highly empirical, as there is no adequate theoretical basis or type of algorithm (as in DGGE [see later]) that would enable researchers to predict the three-dimensional conformation of the single-stranded DNA fragments under specific experimental procedures. Those elements that frequently affect the sensitivity of both methods will be discussed in this section.

In general, most of the single-stranded DNA fragments will have a more compacted conformation at lower temperatures or in the presence of higher salt concentrations. It follows, however, that by increasing the salt concentration, the conductivity is also increased, thus having a rather significant effect on temperature. It is therefore suggested that electrophoresis is performed, as a starting point, at room temperature with 5−10% glycerol or at 4°C without the addition of glycerol or salt. Hayashi and Yandell (1993) have shown that in most fragments shorter than 200 nucleotides, more than 90% of the sequence variations will be detected, and as the size of the PCR fragment gradually increases to 300−350 nucleotides, the number of variants that can be safely detected is reduced to more than 80%. Practical observations suggest that any fragment that exhibits a differential mobility often will migrate very close to the reference fragment. However, the overall fragment number is not always predictable in advance. Any given number of conformations may be supported to a variable extent by the applied electrophoretic parameters. Furthermore, the band intensity is irrelevant to allelic differences, as it is strictly dependent on the different conformations. Therefore SSCP is not a safe method to predict gene-dosage effects. In addition, the simultaneous detection of more than one mutation in a single DNA sample is not easily predictable, as previous data have shown that the electrophoretic pattern may vary considerably within different experiments. In general, although highly reproducible, the mobility of single-stranded DNA conformers cannot be predicted in advance from sequence information. Furthermore, in an attempt to further improve the sensitivity of the SSCP, the RNA−SSCP approach has been developed (Sarkar et al., 1992), where the double-stranded DNA is converted to the corresponding single-stranded RNA by means of one of the two primers that has phage promoter sequences on its 5′ end. Although its sensitivity may rely on the fact that the in vitro transcribed strand has no complementary strand to reanneal with, it still requires sufficient amounts of the in vitro transcribed product to be electrophoresed and easily analyzed, even by ethidium bromide staining (Sarkar et al., 1992).

Due to its numerous advantages, SSCP analysis was (and still is in some cases) the method of choice in several private or public molecular diagnostic laboratories to either interrogate known mutations or scan for known or unknown mutations in a relatively short time. This applies both in basic science as well as in molecular diagnostics. For example, SSCP analysis was first used in screening candidate genes in tumourigenesis (Suzuki et al., 1990; Yap and McGee, 1992). In addition to the utilization of the SSCP methodology as a screening tool, several investigators previously have employed this technique for gene mapping in mouse genes (Beier, 1993) and in fewer cases for the linkage analysis of

human genes (Nishimura et al., 1993; Avramopoulos et al., 1993).

Also, SSCP has been applied on several occasions for the detection of variants in any genomic locus, such as *G6PD* mutation screening (Menounos et al., 2000; Fig. 3.2B) and *HBE* genetic analysis (Papachatzopoulou et al., 2006).

Furthermore, SSCP has been used for molecular typing in clinical microbiology for the detection of nucleotide variations at the groEL gene, in order to differentiate *salmonella* strains both at the interserovar and the intraserovar levels (Nair et al., 2002). Also, Peters et al (1997) reported successful differentiation between the hepatitis C virus (HCV) quasispecies, using F-SSCP analysis of reverse transcription-PCR-generated fragments from HCV RNA, containing hypervariable region I of the HCV genome. Provided that the virus titer is sufficient, these authors showed that F-SSCP is a quick and reliable method for HCV quasispecies analysis. Finally, F-SSCP has been proven to be a particularly useful method in human platelet antigen genotyping (Quintanar et al., 1998).

3.3.2 Heteroduplex Analysis

The principle of HDA is simple and closely related to that of SSCP. In brief, heteroduplexes are formed between different DNA alleles, for example, by mixing wild type and mutant amplified DNA fragments, followed by denaturation at 95°C and slow reannealing to room temperature (White et al., 1992). If the target DNA consists of different alleles already (e.g., a heterozygous case) then heteroduplexes are formed automatically during the amplification step. The result is the formation of two homoduplexes and two heteroduplexes, which are retarded during electrophoresis in native polyacrylamide gels (Fig. 3.2C). There are two types of heteroduplex molecules, depending on the type of the genomic variant, which in turn reflects on their stability. In other words, small deletions or insertions create stable heteroduplexes, termed *bulge type* heteroduplexes, which have been verified by electron microscopy (Wang et al., 1992). On the other hand, single base substitutions form the so-called *bubble type* heteroduplexes, which are much more difficult to visualize, and optimization of the experimental conditions is required to achieve optimal resolution of this type of heteroduplex. A number of molecular diagnostic techniques, based on heteroduplex formation, are reported in the literature, such as enzymatic or chemical cleavage and DGGE, but HDA appears to be the most attractive one, as it can be performed rapidly on short gels without the need for specialized equipment and the use of radioactivity. The most typical example of the use of HDA for mutation screening is the rapid detection of the 3-bp p.F508del deletion in the *CFTR* gene, leading to cystic fibrosis (Fig. 3.2D).

HDA can also be adapted to a fluorescent capillary system, such as the currently available automated DNA sequencers. This is possible by labeling the primers with fluorescent dyes, such as 6-carboxyfluorescein, hexachloro-6-carboxyfluorescein, etc. Like F-SSCP, capillary-based HDA (CE-HDA) also has the advantage of multiplexing, which together with the use of different fluorescent dyes can significantly increase the analysis throughput (Kozlowski and Krzyzosiak, 2001). The use of multicapillary systems, for example, 96-capillary platforms, can also reduce the time needed for processing each sample. For example, up to 10 PCR products can be analyzed in one capillary per run, indicating that the entire analysis of a patient's *BRCA1* and *BRCA2* coding regions can be completed in a single run within 1.5 h (Esteban-Cardeñosa et al., 2004). In brief, CE-HDA is a high-throughput, sufficiently sensitive, and cost-effective methodology and can be easily implemented to offer reliable genetic analysis in molecular diagnostic laboratories with large sample volumes.

The sensitivity of the HDA approach has not been determined to the extent of the SSCP analysis. Rossetti et al. (1995) compared both SSCP and HDA assays for the detection of known mutations in a panel of four genes. Although none of these assays were performed with 100% efficiency, HDA detected slightly more variants than SSCP in the same samples. Also, it has been shown that the performance of HDA can be improved with the gel matrix, in this case mutation detection enhancement (a derivative from HydroLink D5000™; Keen et al., 1991), which has basically made HDA a valuable variant detection technique. Finally, to improve the visualization of the bubble type heteroduplexes, which are much more difficult to visualize than the bulge type heteroduplexes, the Universal Heteroduplex Generator (UHG) was conceived (Wood et al., 1993a,b). The UHG consists of a synthetic DNA fragment, which bears a small (i.e., 2- to 5-bp) deletion, which is amplified by the use of the same oligonucleotide primers as the DNA under study. After amplification, the test amplicon is mixed with the amplified UHG, denatured, and then slowly reannealed, followed by electrophoresis. If no mutation is present in the test DNA, then only a bulge type heteroduplex will be present, slightly retarded compared to the homoduplex. If, however, a single base substitution is also present, then the resulting heteroduplex will have two mismatches: a bulge and a bubble type, which will result in the heteroduplex migrating significantly lower compared to the simple bulge type heteroduplex. The use of UHG has been reported for the detection of known variants within a number of loci, such as von Willebrand disease (Wood et al., 1995), phenylketonuria (Wood et al., 1993a), and for prenatal determination of blood group alleles (Stoerker et al., 1996).

HDA strategies have been successfully designed for the detection of genetic defects in a number of human genes,

such as *CFTR* (Rommens et al., 1990), *NF1* (Shen et al., 1993) and *NF2* (Sainz et al., 1994), *APC* (Paul et al., 1993), *PKU* (Wood et al., 1993a), *BRCA1* and *BRCA2* (using the CE-HDA approach; Kozlowski and Krzyzosiak, 2001; Esteban-Cardeñosa et al., 2004), and so on.

3.3.3 Temperature-Gradient Gel Electrophoresis

TGGE and DGGE are both based on the principle that the electrophoretic mobility of double-stranded DNA fragments is significantly reduced by their partial denaturation. Owing to the sequence dependence of the melting properties of DNA fragments, sequence variations can be detected. Although the sensitivity of TGGE and DGGE in detecting point mutations in genetic disorders and other settings has been reported to be close to 100%, these methods have never become as popular as other mutation detection methods such as SSCP, which may be due to the difficulty in designing adequate PCR primers and setting up the assays.

The method was originally conceived by Myers and coworkers (1985b) for separating DNA fragments differing by single nucleotide substitutions in denaturing gradient gels. The method was based on the notion that the denaturation (melting) of DNA fragments can be regarded as equilibrium for each bp between two distinct states: double helical and a more random state in which bases are neither paired nor stacked on adjacent bases in any orderly way (Myers et al., 1987). The change from the first to the second state is caused by increasing temperature or increasing concentration of denaturing agents.

In the case of single nucleotide substitutions, the replacement of an A:T bp (two hydrogen bonds) by a G:C pair (three hydrogen bonds) will generally be expected to increase the temperature at which the corresponding DNA sequence melts. The context of the nucleotide substitution also plays a role, and substitutions of A:T by T:A pairs, or G:C by C:G pairs, can also affect the temperature at which a DNA sequence dissociates.

Furthermore, a DNA fragment dissociates in a stepwise fashion as the temperature is gradually increased. Dissociation occurs nearly simultaneously in distinct, approximately 50−300-nucleotide-long regions, termed “melting domains.” All nucleotides in a given melting domain dissociate in an all-or-nothing manner within a narrow temperature interval.

The T_m indicates the temperature at which 50% of the individual molecules are dissociated in the given melting domain and 50% are double helical. As indicated earlier, the T_m is strongly dependent on the individual DNA sequence and can be significantly altered by small changes in the DNA sequence, including single nucleotide substitutions. TGGE is based on detecting differences in the electrophoretic mobility between molecules that may differ only at a single position. DNA fragments produced by PCR are subjected to electrophoresis through a linearly increasing gradient of temperature (or a concentration gradient of denaturing agents such as urea and formamide in the case of DGGE). Nucleotide substitutions and other small changes in the DNA sequence are associated with additional bands following TGGE.

The electrophoretic mobility of DNA fragments differs according to whether the fragment is completely double helical, if one or more melting domains has dissociated, or if complete dissociation to two single-stranded molecules has occurred. The electrophoretic mobility of a double helical (nondenatured) DNA fragment is not significantly altered by single nucleotide substitutions within it, but it is primarily dependent on the length and perhaps the curvature of the fragment (Haran et al., 1994). Therefore, assuming that PCR products contain a mixture of two DNA fragments that differ at a single position, as would be the case of a genomic variant in heterozygosity, then both fragments will initially progress through the gel at the same speed. When the molecules reach that point in the gel where the temperature equals their T_m, the molecules will experience a decrease in mobility owing to a transition from a completely double helical (duplex) conformation to a partially denatured one. Dissociation of the first or first few melting domains generally results in a dramatic reduction in the mobility of the DNA fragment, because the fragment takes on a complex, branched conformation. Due to the strong sequence dependence of the T_m, dissociation and consequent retardation of electrophoretic mobility occur at different levels of the temperature gradient associated with bands at different positions in the gel (Myers et al., 1987). In addition to the two homoduplex molecules (wt/wt and mt/mt), two different heteroduplex molecules (wt/mt and mt/wt) can be formed by dissociating and reannealing DNA fragments containing a heterozygous mutation prior to performing TGGE (Fig. 3.3A). Heteroduplex fragments then contain unpaired bases or “bulges” in the otherwise double helical DNA, resulting in a significant reduction in the T_m of the affected melting domain (Ke and Wartell, 1995). The T_m of the two heteroduplex molecules are generally different from one another, so that each heteroduplex is separately visible in the gel. As such, a genomic variant in heterozygosity will thus be visualized by the appearance of four bands: a band representing the normal allele (homoduplex), a band representing the mutant homoduplex that will lie above or underneath the wild type homoduplex band, depending on the effect of the variant on the T_m, and two heteroduplex bands that are always above the homoduplex bands (Fig. 3.3B; Myers et al., 1987). Mutant and wild type homoduplex bands are separated by 2−10 mm in a typical polyacrylamide gel, and the

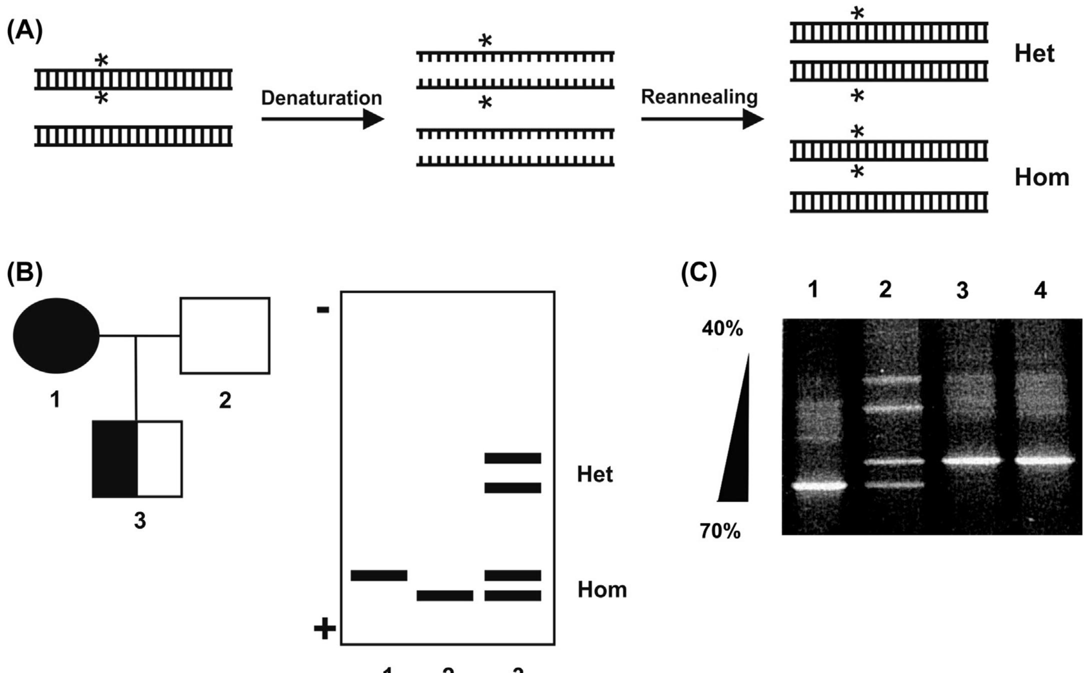


FIGURE 3.3 (A) Mechanism of heteroduplex formation. In the case of heterozygous point mutations in genetic disorders (indicated with an asterisk), the polymerase chain reaction produces two alleles differing only at the position of the point mutation. Wild type and mutant molecules are present at an approximately 1:1 ratio. Denaturation followed by reannealing of these molecules produces both wild type and mutant homoduplex molecules (Hom) as well as two heteroduplex molecules, consisting of a wild type and a mutant strand (Het). (B) Parallel temperature- and denaturing-gradient gel electrophoresis (DGGE). Mutation screening is generally performed with the temperature or denaturing gradient parallel to the direction of electrophoresis. In this example, the results of electrophoresis from top to bottom for a hypothetical family segregating an autosomal recessive disorder are shown. Case 2 is normal, carrying only the wild type allele; Case 3 is heterozygous for a point mutation, resulting in the appearance of an additional homoduplex band as well as two additional heteroduplex bands (Het). Case 1, who is homozygous for the mutation, shows just the mutant homoduplex band (Hom). (C) Mutation detection using DGGE analysis. Screening for the GC-158 C > T polymorphism in the promoter region of the human *HBG2* (GC-globin) gene. Lanes 3 and 4 correspond to homozygous samples for that polymorphism. Lane 2 corresponds to a heterozygous sample, and lane 1 corresponds to a sample that does not carry this polymorphism on either of the two alleles. *Photo courtesy of George P. Patrinos.*

heteroduplex bands are often 3 or more cm above the homoduplex bands.

An important issue in both TGGE (and DGGE) is that genomic variants are detectable only in the melting domain(s) with the lowest T_m. As such, it becomes problematic to visualize genomic variants that are present in DNA fragments with multiple melting domains with different T_m, since once the DNA fragments reach the temperature at which the first melting domain dissociates, the mobility of the fragment is greatly reduced so that it may not reach temperatures relevant for the higher T_m domains under the conditions of the experiment. In other words, only mutations within the lowest T_m domain can be reliably detected by TGGE or DGGE (Myers et al., 1987).

To overcome this bottleneck, the concept of the GC-clamp has been introduced. The GC-clamp is a 135-bp-long GC-rich sequence that is attached to the fragment to be analyzed in the amplification step. Without the GC-clamp, only variants present in the lowest melting domain (the one with the lower T_m) could be visualized in the gel. However, owing to its high GC content, the GC-clamp has a significantly higher T_m than most naturally occurring sequences, and as such, the addition of the GC-clamp significantly alters the melting properties of any sequence, allowing the detection of any variant in this fragment (Myers et al., 1985a). It has been shown that the addition of a 40-nucleotide-long G + C-rich sequence to one of the two PCR primers, a GC-clamp can be conveniently added to any DNA fragment produced by PCR (Sheffield et al., 1989). It is also possible to use a universal GC-clamp that is incorporated into amplified DNA fragments during PCR, thereby avoiding the expense of synthesizing long primers (Top, 1992). Psoralen-modified PCR primers can be alternatively used to GC-clamps. One of the two PCR primers is 5′ modified by 5-(⌉-hexyloxy)-psoralen. The 5′ terminus of the primer should have two adenosine residues; if the natural sequence does not have AA, this sequence should be appended to the specific DNA sequence of the primer. Psoralens are bifunctional photoreagents that can form covalent bonds with pyrimidine bases (especially thymidine). If intercalated at 5′-TpT in double helical DNA (this will be the complementary sequence of the 3′ terminus of

the other strand following PCR), psoralen forms a covalent bond with thymidine after photoinduction (Costes et al., 1993b). Photoinduction can be performed by exposing the PCR products to a source of ultraviolet (UV) light (365) for 5−15 min, which can be conveniently done in the original PCR tubes or a 96-well plate. Psoralen clamping is sometimes preferred over GC-clamping because the PCR is often easier to optimize, and bipolar clamping is possible if necessary. Psoralen modification of primers is available from many commercial oligonucleotide sources.

To optimize the gel running conditions, a computer program, such as Melt94, can be used, but it would still be necessary to perform some preliminary experiments to determine the optimal electrophoresis conditions and running times and to confirm that the optimal denaturing gradient has been chosen to ensure that bands get well-separated. The Melt87 program calculates the T_m for each bp in the DNA fragment, that is, the temperature at which 50% of the individual molecules are double helical and 50% of the molecules are in a fully melted state. The results of such a calculation are termed the "melting map," where DNA fragments are shown to be typically divided into distinct melting domains of about 50−300 bp in length, in which all bps have nearly identical T_m. The melting map demonstrates the lowest melting domain in the DNA fragment; as mentioned earlier, only genomic variants in this region will be visible by DGGE or TGGE analysis. Similarly, to choose the optimal denaturant concentration range, specific software (SQHTX, Lerman and Silverstein, 1987) can be used, which calculates the difference in temperature at which the homoduplex and the heteroduplex molecules partially melt. Subsequently, the optimization process begins with a perpendicular TGGE experiment, in which electrophoresis is performed perpendicularly to the temperature gradient. Perpendicular TGGE is run with a gradient of 20−60°C, which will be adequate for the vast majority of PCR fragments.

3.3.4 Denaturing Gradient Gel Electrophoresis

DGGE is a form of polyacrylamide gel electrophoresis in which a double-stranded DNA fragment migrates into a gradient of linearly increasing denaturing conditions, instead of temperature, which is functionally equivalent to the temperature gradient of TGGE. The denaturants used are heat (a constant temperature of generally 60°C) and a fixed ratio of formamide (ranging from 0 to 40%) and urea (ranging from 0 to 7 M). The temperature of 60°C was empirically chosen to exceed the T_m of an AT-rich DNA fragment in the absence of a denaturant. Higher temperatures (e.g., 75°C) can be employed for very GC-rich DNA sequences. To achieve a uniform temperature distribution the electrophoresis unit is submerged into a circulating water bath.

DGGE, like TGGE, has been used to investigate a large number of disease genes. Due to the relative ease of detecting genomic variants in heterozygosity, mainly owing to the occurrence of the heteroduplexes, TGGE and DGGE have been particularly useful for disorders characterized by heterozygous variants or frequent de novo mutations (reviewed in Fodde and Losekoot, 1994). In light of the effort involved in designing primers and optimizing conditions, TGGE or DGGE is generally reserved for situations when large numbers of samples are to be screened for mutations.

Most mutation screening protocols involve the simultaneous analysis of 24 or more samples on one parallel TGGE/DGGE. In general, altered band patterns are easy to spot. The classical appearance of heterozygous mutations (Fig. 3.3C) is due to the appearance of three additional bands. With some mutations, only one or two additional bands are seen. In general, without GC-clamping, Myers et al., (1985b) have shown a detection rate of nearly 40% of the sequence variants in a DNA fragment up to 500 bp, but the addition of psoralen or GC-clamps can significantly improve the detection rate, in most cases up to 100%. The sensitivity of TGGE/DGGE for detecting known variants is generally reported to be almost 100%, as in the case of the *CFTR* gene, where DGGE detected 201 of 201 known variants (Macek et al., 1997). Also, TGGE and DGGE have been shown to be very sensitive in the detection of mutations in situations where the mutation sequence is present in proportions less than 50%, such as in the case of heteroplasmy in mitochondrial disorders with heteroplasmic proportions of as low as 1% (Tully et al., 2000), as well as in testing for residual disease in cancer (Ahnhudt et al., 2001; Alkan et al., 2001).

There have also been several variations of the main DGGE and TGGE themes, such as *broad range DGGE*, where a single gel and a single set of conditions are used to screen all of the exons of one gene (Guldberg and Guttler, 1994; Hayes et al., 1999); *multiplex DGGE*, where several exons are simultaneously analyzed in one DGGE gel (Costes et al., 1993a); *genomic DGGE*, where genomic DNA is digested with a restriction enzyme, electrophoresed by DGGE, transferred to a nylon membrane, and hybridized to a unique DNA probe (Borresen et al., 1988); *constant DGGE (cDGGE)*, where gels contain constant concentrations of denaturants, allowing for an increased resolution of mutant fragments since they will constantly migrate with a different electrophoretic mobility through the whole length of the gel (Hovig et al., 1991); *constant denaturant capillary electrophoresis*, where the separation of DNA fragments is achieved by the differential velocity of partly melted DNA in a medium with a uniform denaturant concentration (also available in an array-based

format); *chip-based temperature gradient capillary electrophoresis* (Zhang et al., 2007); **temporal temperature gradient gel electrophoresis**, where a constant concentration of urea or formamide is used, as in cDGGE, but the temperature during the run is gradually increased (Yoshino et al., 1991; Wiese et al., 1995); *microtemperature-gradient gel electrophoresis*, where a minimized gel (20 × 20 × 0.5 mm) is used to minimize the amount of DNA required and the running time (approximately 12 min at 100 V, 10 mA; Biyani and Nishigaki, 2001; Tominaga, 2007); *double-gradient, denaturing gradient gel electrophoresis*, where in addition to the chemical denaturing gradient (formamide and urea) a sieving gradient (e.g., 6–12% polyacrylamide gradient) is used (Cremonesi et al., 1997); *two-dimensional DNA fingerprinting/two-dimensional gene scanning*, combining the size separation of DNA fragments in the first dimension with their sequence-specific separation through DGGE in the second dimension; *denaturing HPLC*, using an ion-pair chromatography separation principle combined with a precise control of the column temperature and optimized mobile phase gradient for the separation of mutant DNA molecules (reviewed in Xiao and Oefner, 2001); and *cycling gradient capillary electrophoresis*, where DNA sequence variants are detected based on their differential migration in a polymer-filled capillary system, allowing multiple samples to be analyzed into the same capillary (or set of capillaries) separated by predefined time intervals of partial electrophoresis (e.g., a 96-capillary system is able to screen over 15,000 samples in 24 h; Minarik et al., 2003).

There have been several diagnostic applications of both DGGE and TGGE (Table 3.1). However, the gradual use of other high-throughput techniques, such as microarrays and next-generation sequencing, has led to the displacement of TGGE and DGGE.

3.4 CONCLUSIONS

All of the techniques described earlier have been extremely popular and constitute the basis on which several diagnostic laboratories are operating. Allele-specific mutation detection methods, particularly PCR–ARMS and PCR–ASO, are feasible and quite straightforward with a comparably high degree of accuracy and specificity, although they are only suitable for low-throughput laboratories.

TABLE 3.1 Indicative Examples of Denaturing Gradient Gel Electrophoresis Applications for Molecular Diagnostics

Application	Gene	References
Screening of genes in the human nuclear genome	*COL1A2*	Borresen et al. (1988)
	SERPINA1	Hayes (2003)
	HBG1/HBG2	Patrinos et al. (1998, 2001)
	TP53	Pignon et al. (1994)
	FBN1	Tiecke et al. (2001), Katzke et al. (2002), and Robinson et al. (2002)
	NF1	Peters et al. (1999) and Fahsold et al., 2000
	DMD	Hofstra et al. (2004)
	HBD	Papadakis et al. (1997)
	HBB	Losekoot et al. (1990)
	MEN1	Balogh et al. (2004)
	Human Leukocyte Antigen typing	Uhrberg et al. (1994) and Knapp (2005)
	Single nucleotide polymorphism discovery	Hsia et al. (2005) and Maher et al. (2006)
Mitochondrial genome	Mitochondrial variants	Hanekamp et al. (1996) and Chen et al. (1999)
Molecular microbiology	Bacterial analysis	Muyzer and Smalla (1998), van Elsas et al. (2002), Deng et al. (2008), and Ercolini (2004)
	Microbial genome profiling	Watanabe et al. (2002)
	Viral genome screening	Lu et al. (2002) and Motta et al. (2002)
Plant molecular biodiversity		Gomes et al. (2003) and Nikolcheva et al. (2003)

Similarly, SSCP and HDA, being simple, reliable, and sensitive methods for the detection of nucleotide sequence changes in genomic loci, are broadly used for the DNA analysis of human cancers and other genetic disorders. TGGE, DGGE, and the closely related methods provide a very high sensitivity and are relatively easy and cheap alternatives to perform once the assays have been designed and optimized. The main advantages of all of these approaches are a high detection rate and specificity and improved heterozygote detection. The main disadvantages are the limitations of the PCR fragment length (300−500 nucleotides) and the difficulties of analyzing GC-rich fragments (in the case of DGGE/TGGE).

Although these techniques once dominated in molecular diagnostics laboratories, they gradually tend to be abandoned. Nevertheless, it is noteworthy that some of these methodologies, such as allele-specific mutation detection, set the standards for the development of high-throughput systems, such as microarrays for the purpose of large-scale SNP genotyping that concern genome scans and DNA diagnostics of genetic, acquired, and infectious diseases (see also Chapter 18). Initial efforts, attempted for *HBB* mutation screenings in an array format (Foglieni et al., 2004), have paved the path for the development of today's high-throughput microarrays.

REFERENCES

Ahnhudt, C., Muche, J.M., Dijkstal, K., Sterry, W., Lukowsky, A., 2001. An approach to the sensitivity of temperature-gradient gel electrophoresis in the detection of clonally expanded T-cells in cutaneous T-cell lymphoma. Electrophoresis 22, 33−38.

Alkan, S., Cosar, E., Ergin, M., Hsi, E., 2001. Detection of T-cell receptor-gamma gene rearrangement in lymphoproliferative disorders by temperature gradient gel electrophoresis. Arch. Pathol. Lab. Med. 125, 202−207.

Alves, A.M., Carr, F.J., 1988. Dot blot detection of point mutations with adjacently hybridising synthetic oligonucleotide probes. Nucleic Acids Res. 16, 8723.

Avramopoulos, D., Cox, T., Kraus, J.P., Chakravarti, A., Antonarakis, S.E., 1993. Linkage mapping of the cystathionine beta-synthase (CBS) gene on human chromosome 21 using a DNA polymorphism in the 3′ untranslated region. Hum. Genet. 90, 566−568.

Balogh, K., Patócs, A., Majnik, J., Rácz, K., Hunyady, L., 2004. Genetic screening methods for the detection of mutations responsible for multiple endocrine neoplasia type 1. Mol. Genet. Metab. 83, 74−81.

Barany, F., 1991. Genetic disease detection and DNA amplification using cloned thermostable ligase. Proc. Natl. Acad. Sci. U.S.A. 88, 189−193.

Beier, D.R., 1993. Single-strand conformation polymorphism (SSCP) analysis as a tool for genetic mapping. Mamm. Genome 4, 627−631.

Biyani, M., Nishigaki, K., 2001. Hundredfold productivity of genome analysis by introduction of microtemperature-gradient gel electrophoresis. Electrophoresis 22, 23−28.

Borresen, A.L., Hovig, E., Brogger, A., 1988. Detection of base mutations in genomic DNA using denaturing gradient gel electrophoresis (DGGE) followed by transfer and hybridization with gene-specific probes. Mutat. Res. 202, 77−83.

Cai, S.P., Wall, J., Kan, Y.W., Chehab, F.F., 1994. Reverse dot blot probes for the screening of b-thalassemia mutations in Asians and American Blacks. Hum. Mutat. 3, 59−63.

Chan, V., Yam, I., Chen, F.E., Chan, T.K., 1999. A reverse dot-blot for rapid detection of non-deletion a thalassaemia. Br. J. Haematol. 104, 513−515.

Chehab, F.F., Wall, J., 1992. Detection of multiple cystic fibrosis mutations by reverse dot blot hybridization: a technology for carrier screening. Hum. Genet. 89, 163−168.

Chehab, F.F., 1993. Molecular diagnostics: past, present, and future. Hum. Mutat. 2, 331−337.

Chen, T.J., Boles, R.G., Wong, L.J., 1999. Detection of mitochondrial DNA mutations by temporal temperature gradient gel electrophoresis. Clin. Chem. 45, 1162−1167.

Conner, B.J., Reyes, A.A., Morin, C., Itakurs, K., Teplitz, R.L., Wallace, R.B., 1983. Detection of sickle cell S-globin allele by hybridization with synthetic oligonucleotides. Proc. Natl. Acad. Sci. U.S.A. 80, 278−282.

Costes, B., Fanen, P., Goossens, M., Ghanem, N., 1993a. A rapid, efficient, and sensitive assay for simultaneous detection of multiple cystic fibrosis mutations. Hum. Mutat. 2, 185−191.

Costes, B., Girodon, E., Ghanem, N., Chassignol, M., Thuong, N.T., Dupret, D., Goossens, M., 1993b. Psoralen-modified oligonucleotide primers improve detection of mutations by denaturing gradient gel electrophoresis and provide an alternative to GC-clamping. Hum. Mol. Genet. 2, 393−397.

Cremonesi, L., Firpo, S., Ferrari, M., Righetti, P.G., Gelfi, C., 1997. Double-gradient DGGE for optimized detection of DNA point mutations. Biotechniques 22, 326−330.

Day, D.J., Speiser, P.W., White, P.C., Barany, F., 1995a. Detection of steroid 21-hydroxylase alleles using gene-specific PCR and a multiplexed ligation detection reaction. Genomics 29, 152−162.

Day, I.N., Whittall, R., Gudnason, V., Humphries, S.E., 1995b. Dried template DNA, dried PCR oligonucleotides and mailing in 96-well plates: LDL receptor gene mutation screening. Biotechniques 18, 981−984.

Day, N.S., Tadin, M., Christiano, A.M., Lanzano, P., Piomelli, S., Brown, S., 2002. Rapid prenatal diagnosis of sickle cell diseases using oligonucleotide ligation assay coupled with laser-induced capillary fluorescence detection. Prenat. Diagn. 22, 686−691.

Deng, W., Xi, D., Mao, H., Wanapat, M., 2008. The use of molecular techniques based on ribosomal RNA and DNA for rumen microbial ecosystem studies: a review. Mol. Biol. Rep. 35, 265−274.

Ercolini, D., 2004. PCR-DGGE fingerprinting: novel strategies for detection of microbes in food. J. Microbiol. Methods 56 (3), 297−314.

Esteban-Cardeñosa, E., Duran, M., Infante, M., Velasco, E., Miner, C., 2004. High-throughput mutation detection method to scan BRCA1 and BRCA2 based on heteroduplex analysis by capillary array electrophoresis. Clin. Chem. 50, 313−320.

Fahsold, R., Hoffmeyer, S., Mischung, C., Gille, C., Ehlers, C., Kucukceylan, N., Abdel-Nour, M., Gewies, A., Peters, H., Kaufmann, D., Buske, A., Tinschert, S., Nurnberg, P., 2000. Minor lesion mutational spectrum of the entire NF1 gene does not explain its high mutability but points to a functional domain upstream of the GAP-related domain. Am. J. Hum. Genet. 66, 790−818.

Fodde, R., Losekoot, M., 1994. Mutation detection by denaturing gradient gel electrophoresis (DGGE). Hum. Mutat. 3, 83−94.

Foglieni, B., Cremonesi, L., Travi, M., Ravani, A., Giambona, A., Rosatelli, M.C., Perra, C., Fortina, P., Ferrari, M., 2004. Beta-

thalassemia microelectronic chip: a fast and accurate method for mutation detection. Clin. Chem. 50, 73−79.

Foglietta, E., Bianco, I., Maggio, A., Giambona, A., 2003. Rapid detection of six common Mediterranean and three non-Mediterranean a-thalassemia point mutations by reverse dotblot analysis. Am. J. Hematol. 74, 191−195.

Fox, D.H., Huang, C.K., Du, J., Chang, T.Y., Pan, Q., 2007. Profound inhibition of the PCR step of CF V3 multiplex PCR/OLA assay by the use of UV-irradiated plastic reaction tubes. Diagn. Mol. Pathol. 16, 121−123.

Gold, B., 2003. Origin and utility of the reverse dot-blot. Expert Rev. Mol. Diagn. 3, 143−152.

Gomes, N.C., Fagbola, O., Costa, R., Rumjanek, N.G., Buchner, A., Mendona-Hagler, L., Smalla, K., 2003. Dynamics of fungal communities in bulk and maize rhizosphere soil in the tropics. Appl Environ. Microbiol. 69 (7), 3758−3766. Erratum in: Appl. Environ. Microbiol. 69 (9), 5737.

Grossman, P.D., Bloch, W., Brinson, E., Chang, C.C., Eggerding, F.A., Fung, S., Iovannisci, D.M., Woo, S., Winn-Deen, E.S., Iovannisci, D.A., 1994. High-density multiplex detection of nucleic acid sequences: oligonucleotide ligation assay and sequence-coded separation. Nucleic Acids Res. 22, 4527−4534.

Guldberg, P., Guttler, F., 1994. 'Broad-range' DGGE for single-step mutation scanning of entire genes: application to human phenylalanine hydroxylase gene. Nucleic Acids Res. 22, 880−881.

Hanekamp, J.S., Thilly, W.G., Chaudhry, M.A., 1996. Screening for human mitochondrial DNA polymorphisms with denaturing gradient gel electrophoresis. Hum. Genet. 98, 243−245.

Hansen, T.S., Petersen, N.E., Iitia, A., Blaabjerg, O., Hyltoft- Petersen, P., Horder, M., 1995. Robust nonradioactive oligonucleotide ligation assay to detect a common point mutation in the *CYP2D6* gene causing abnormal drug metabolism. Clin. Chem. 41, 413−418.

Haran, T.E., Kahn, J.D., Crothers, D.M., 1994. Sequence elements responsible for DNA curvature. J. Mol. Biol. 244, 135−143.

Hayashi, K., Yandell, D.W., 1993. How sensitive is PCRSSCP? Hum. Mutat. 2, 338−346.

Hayashi, K., 1991. PCR-SSCP: a simple and sensitive method for detection of mutations in the genomic DNA. PCR Methods Appl. 1, 34−38.

Hayes, V.M., Wu, Y., Osinga, J., Mulder, I.M., van der Vlies, P., Elfferich, P., Buys, C.H., Hofstra, R.M., 1999. Improvements in gel composition and electrophoretic conditions for broad-range mutation analysis by denaturing gradient gel electrophoresis. Nucleic Acids Res. 27, e29.

Hayes, V.M., 2003. Genetic diversity of the alpha-1-antitrypsin gene in Africans identified using a novel genotyping assay. Hum. Mutat. 22, 59−66.

Hofstra, R.M., Mulder, I.M., Vossen, R., de Koning-Gans, P.A., Kraak, M., Ginjaar, I.B., van der Hout, A.H., Bakker, E., Buys, C.H., van Ommen, G.J., van Essen, A.J., den Dunnen, J.T., 2004. DGGE-based whole-gene mutation scanning of the dystrophin gene in Duchenne and Becker muscular dystrophy patients. Hum. Mutat. 23, 57−66.

Hovig, E., Smith-Sorensen, B., Brogger, A., Borresen, A.L., 1991. Constant denaturant gel electrophoresis, a modification of denaturing gradient gel electrophoresis, in mutation detection. Mutat. Res. 262, 63−71.

Hsia, A.P., Wen, T.J., Chen, H.D., Liu, Z., Yandeau-Nelson, M.D., Wei, Y., Guo, L., Schnable, P.S., 2005. Temperature gradient capillary electrophoresis (TGCE) − a tool for the high-throughput discovery and mapping of SNPs and IDPs. Theor. Appl. Genet. 111, 218−225.

Humphries, S.E., Gudnason, V., Whittall, R., Day, I.N., 1997. Single-strand conformation polymorphism analysis with high throughput modifications, and its use in mutation detection in familial hypercholesterolemia. International Federation of Clinical Chemistry Scientific Division: Committee on Molecular Biology Techniques. Clin. Chem. 43, 427−435.

Iannone, M.A., Taylor, J.D., Chen, J., Li, M.S., Rivers, P., Slentz-Kesler, K.A., Weiner, M.P., 2000. Multiplexed single nucleotide polymorphism genotyping by oligonucleotide ligation and flow cytometry. Cytometry 39, 131−140.

Katzke, S., Booms, P., Tiecke, F., Palz, M., Pletschacher, A., Turkmen, S., Neumann, L.M., Pregla, R., Leitner, C., Schramm, C., Lorenz, P., Hagemeier, C., Fuchs, J., Skovby, F., Rosenberg, T., Robinson, P.N., 2002. TGGE screening of the entire FBN1 coding sequence in 126 individuals with Marfan syndrome and related fibrillinopathies. Hum. Mutat. 20, 197−208.

Ke, S.H., Wartell, R.M., 1995. Influence of neighboring base pairs on the stability of single base bulges and base pairs in a DNA fragment. Biochemistry 34, 4593−4600.

Keen, J., Lester, D., Inglehearn, C., Curtis, A., Bhattacharya, S., 1991. Rapid detection of single base mismatches as heteroduplexes on Hydrolink gels. Trends Genet. 7, 5.

Knapp, L.A., 2005. Denaturing gradient gel electrophoresis and its use in the detection of major histocompatibility complex polymorphism. Tissue Antigens 5, 211−219.

Kozlowski, P., Krzyzosiak, W.J., 2001. Combined SSCP/duplex analysis by capillary electrophoresis for more efficient mutation detection. Nucleic Acids Res. 29, E71.

Landegren, U., Kaiser, R., Sanders, J., Hood, L., 1988. Aligase-mediated gene detection technique. Science 241, 1077−1080.

Lerman, L.S., Silverstein, K., 1987. Computational simulation of DNA melting and its application to denaturing gradient gel electrophoresis. Methods Enzymol. 155, 482−501.

Liu, J., Zhao, Z., Sun, M., Chen, K., Yuan, W., Jiang, G., 2016. The sensitive detection of telomerase reverse transcriptase promoter mutation by amplification refractory mutation system-PCR. Genet. Test. Mol. Biomarkers 20, 90−93.

Losekoot, M., Fodde, R., Harteveld, C.L., van Heeren, H., Giordano, P.C., Bernini, L.F., 1990. Denaturing gradient gel electrophoresis and direct sequencing of PCR amplified genomic DNA: a rapid and reliable diagnostic approach to beta thalassaemia. Br. J. Haematol. 76, 269−274.

Lu, Q., Hwang, Y.T., Hwang, C.B., 2002. Mutation spectra of herpes simplex virus type 1 thymidine kinase mutants. J. Virol. 76 (11), 5822−5828.

Macek Jr., M., Mercier, B., Mackova, A., Miller, P.W., Hamosh, A., Ferec, C., Cutting, G.R., 1997. Sensitivity of the denaturing gradient gel electrophoresis technique in detection of known mutations and novel Asian mutations in the CFTR gene. Hum. Mutat. 9, 136−147.

Maher, P.M., Chou, H.H., Hahn, E., Wen, T.J., Schnable, P.S., 2006. GRAMA: genetic mapping analysis of temperature gradient capillary electrophoresis data. Theor. Appl. Genet. 113, 156−162.

Makino, R., Yazyu, H., Kishimoto, Y., Sekiya, T., Hayashi, K., 1992. F-SSCP: fluorescence-based polymerase chain reaction-single-strand conformation polymorphism (PCR-SSCP) analysis. PCR Methods Appl. 2, 10−13.

Makowski, G.S., Nadeau, F.L., Hopfer, S.M., 2003. Single tube multiplex PCR detection of 27 cystic fibrosis mutations and 4 polymorphisms

using neonatal blood samples collected on Guthrie cards. Ann. Clin. Lab. Sci. 33, 243–250.

Martinelli, R.A., Arruda, J.C., Dwivedi, P., 1996. Chemiluminescent hybridization-ligation assays for delta F508 and delta I507 cystic fibrosis mutations. Clin. Chem. 42, 14–18.

Medikare, V., Ali, A., Ananthapur, V., Deendayal, M., Nallari, P., 2016. Susceptibility risk alleles of -238G/A, -308G/A and -1031T/C promoter polymorphisms of TNF-α gene to uterine leiomyomas. Recent Adv. DNA Gene Seq. 9, 65–71.

Menounos, P., Zervas, C., Garinis, G., Doukas, C., Kolokithopoulos, D., Tegos, C., Patrinos, G.P., 2000. Molecular heterogeneity of the glucose-6-phosphate dehydrogenase deficiency in the Hellenic population. Hum. Hered. 50, 237–241.

Minarik, M., Minarikova, L., Bjorheim, J., Ekstrom, P.O., 2003. Cycling gradient capillary electrophoresis: a low-cost tool for high-throughput analysis of genetic variations. Electrophoresis 24, 1716–1722.

Motta, F.C., Rosado, A.S., Couceiro, J.N., 2002. Standardization of denaturing gradient gel electrophoresis for mutant screening of influenza A (H3N2) virus samples. J. Virol. Methods 101 (1–2), 105–115.

Mullis, K.B., Faloona, F.A., 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol 155, 335–350.

Muyzer, G., Smalla, K., 1998. Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. Ant. Van Leeuwenhoek 73, 127–141.

Myers, R.M., Fischer, S.G., Lerman, L.S., Maniatis, T., 1985a. Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. Nucleic Acids Res. 13, 3131–3145.

Myers, R.M., Lumelsky, N., Lerman, L.S., Maniatis, T., 1985b. Detection of single base substitutions in total genomic DNA. Nature 313, 495–498.

Myers, R.M., Maniatis, T., Lerman, L.S., 1987. Detection and localization of single base changes by denaturing gradient gel electrophoresis. Methods Enzymol. 155, 501–527.

Nair, S., Lin, T.K., Pang, T., Altwegg, M., 2002. Characterization of *Salmonella* serovars by PCR-single-strand conformation polymorphism analysis. J. Clin. Microbiol. 40, 2346–2351.

Newton, C.R., Graham, A., Heptinstall, L.E., Powell, S.J., Summers, C., Kalsheker, N., Smith, J.C., Markham, A.F., 1989. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res. 17, 2503–2516.

Nichols, W.C., Liepieks, J.J., McKusick, V.A., Benson, M.D., 1989. Direct sequencing of the gene for Maryland/German familial amyloidotic polyneuropathy type II and genotyping by allele-specific enzymatic amplification. Genomics 5, 535–540.

Nickerson, D.A., Kaiser, R., Lappin, S., Stewart, J., Hood, L., Landegren, U., 1990. Automated DNA diagnostics using an ELISA-based oligonucleotide ligation assay. Proc. Natl. Acad. Sci. U.S.A. 87, 8923–8927.

Nishimura, D.Y., Purchio, A.F., Murray, J.C., 1993. Linkage localization of TGFB2 and the human homeobox gene HLX1 to chromosome 1q. Genomics 15, 357–364.

Nowaczyk, M.J., Nakamura, L.M., Eng, B., Porter, F.D., Waye, J.S., 2001. Frequency and ethnic distribution of the common DHCR7 mutation in Smith-Lemli-Opitz syndrome. Am. J. Med. Genet. 102, 383–386.

Okayama, H., Curiel, D.T., Brantly, M.L., Holmes, M.D., Crystal, R.D., 1989. Rapid nonradioactive detection of mutations in the human genome by allele-specific amplification. J. Lab. Clin. Med. 114, 105–113.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., Sekiya, T., 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc. Natl. Acad. Sci. U.S.A. 86, 2766–2770.

Orkin, S.H., Markham, A.F., Kazazian Jr., H.H., 1983. Direct detection of the common Mediterranean beta-thalassemia gene with synthetic DNA probes: an alternate approach for prenatal diagnosis. J. Clin. Invest. 71, 775–779.

Papachatzopoulou, A., Menounos, P.G., Kolonelou, C., Patrinos, G.P., 2006. Mutation screening in the human epsilon-globin gene using single-strand conformation polymorphism analysis. Am. J. Hematol. 81, 136–138.

Papadakis, M., Papapanagiotou, E., Loutradi-Anagnostou, A., 1997. Scanning method to identify the molecular heterogeneity of delta-globin gene especially in delta-thalassemias: detection of three novel substitutions in the promoter region of the gene. Hum. Mutat. 9, 465–472.

Patrinos, G.P., Kollia, P., Loutradi-Anagnostou, A., Loukopoulos, D., Papadakis, M.N., 1998. The Cretan type of non-deletional hereditary persistence of fetal hemoglobin [Agamma-158C->T] results from two independent gene conversion events. Hum. Genet. 102, 629–634.

Patrinos, G.P., Kollia, P., Papapanagiotou, E., Loutradi-Anagnostou, A., Loukopoulos, D., Papadakis, M.N., 2001. Agamma-haplotypes: a new group of genetic markers for thalassemic mutations inside the 5' regulatory region of the human Agamma-globin gene. Am. J. Hematol. 66, 99–104.

Patrinos, G.P., Kollia, P., Papadakis, M.N., 2005. Molecular diagnosis of thalassemia and hemoglobinopathies: from RFLP analysis to microarrays. Hum. Mutat. 26, 399–412.

Paul, P., Letteboer, T., Gelbert, L., Groden, J., White, R., Coppes, M.J., 1993. Identical APC exon 15 mutations result in a variable phenotype in familial adenomatous polyposis. Hum. Mol. Genet. 2, 925–931.

Peters, T., Schlayer, H.J., Hiller, B., Rösler, B., Blum, H., Rasenack, J., 1997. Quasispecies analysis in hepatitis C virus infection by fluorescent single strand conformation polymorphism. J. Virol. Methods 64, 95–102.

Peters, H., Hess, D., Fahsold, R., Schulke, M., 1999. A novel mutation L1425P in the GAP-region of the NF1 gene detected by temperature gradient gel electrophoresis (TGGE). Mutation in brief no. 230. Online. Hum. Mutat. 13, 337.

Pignon, J.M., Vinatier, I., Fanen, P., Jonveaux, P., Tournilhac, O., Imbert, M., Rochant, H., Goossens, M., 1994. Exhaustive analysis of the P53 gene coding sequence by denaturing gradient gel electrophoresis: application to the detection of point mutations in acute leukemias. Hum. Mutat. 3, 126–132.

Pirastu, M., Kan, Y.W., Cao, A., Conner, B.J., Teplitz, R.L., Wallace, R.B., 1983. Prenatal diagnosis of b-thalassemia: detection of a single nucleotide mutation in DNA. N. Engl. J. Med. 309, 284–287.

Pritchard, C.E., Southern, E.M., 1997. Effects of base mismatches on joining of short oligodeoxynucleotides by DNA ligases. Nucleic Acids Res. 25, 3403–3407.

Quintanar, A., Jallu, V., Legros, Y., Kaplan, C., 1998. Human platelet antigen genotyping using a fluorescent SSCP technique with an automatic sequencer. Br. J. Haematol. 103, 437−444.

Robinson, P.N., Booms, P., Katzke, S., Ladewig, M., Neumann, L., Palz, M., Pregla, R., Tiecke, F., Rosenberg, T., 2002. Mutations of FBN1 and genotype-phenotype correlations in Marfan syndrome and related fibrillinopathies. Hum. Mutat. 20, 153−161.

Rommens, J., Kerem, B.S., Greer, W., Chang, P., Tsui, L.C., Ray, P., 1990. Rapid nonradioactive detection of the major cystic fibrosis mutation. Am. J. Hum. Genet. 46, 395−396.

Rossetti, S., Corra, S., Biasi, M.O., Turco, A.E., Pignatti, P.F., 1995. Comparison of heteroduplex and single-strand conformation analyses, followed by ethidium fluorescence visualization, for the detection of mutations in four human genes. Mol. Cell. Probes 9, 195−200.

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., Arnheim, N., 1985. Enzymatic amplification of b globin sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230, 1350−1354.

Saiki, R.K., Bugawan, T.L., Horn, G.T., Mullis, K.B., Erlich, H.A., 1986. Analysis of enzymatically amplified b-globin and HLA-DQa DNA with allele-specific oligonucleotide probes. Nature 324, 163−166.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., Erlich, H.A., 1988a. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239, 487−491.

Saiki, R.K., Chang, C.-A., Levenson, C.H., Warren, T.C., Boehm, C.D., Kazazian Jr., H.H., Erlich, H.A., 1988b. Diagnosis of sickle cell anemia and b-thalassemia with enzymatically amplified DNA and nonradioactive allele-specific oligonucleotide probes. N. Engl. J. Med. 319, 537−541.

Saiki, R.K., Walsh, P.S., Levenson, C.H., Erlich, H.A., 1989. Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. Proc. Natl. Acad. Sci. U.S.A. 86, 6230−6234.

Sainz, J., Huynh, D.P., Figueroa, K., Ragge, N.K., Baser, M.E., Pulst, S.M., 1994. Mutations of the neurofibromatosis Ch06.qxd 4/9/05 3:40 PM Page 64 type 2 gene and lack of the gene product in vestibular schwannomas. Hum. Mol. Genet. 3, 885−891.

Sarkar, G., Yoon, H.S., Sommer, S.S., 1992. Screening for mutations by RNA single-strand conformation polymorphism (rSSCP): comparison with DNA-SSCP. Nucleic Acids Res. 20, 871−878.

Sheffield, V.C., Cox, D.R., Lerman, L.S., Myers, R.M., 1989. Attachment of a 40-base-pair G + C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. Proc. Natl. Acad. Sci. U.S.A. 86, 232−236.

Shen, M.H., Harper, P.S., Upadhyaya, M., 1993. Neurofibromatosis type 1 (NF1): the search for mutations by PCRheteroduplex analysis on Hydrolink gels. Hum. Mol. Genet. 2, 1861−1864.

Sommer, S.S., Cassady, J.D., Sobell, J.L., Bottema, C.D., 1989. 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.

Stoerker, J., Hurwitz, C., Rose, N.C., Silberstein, L.E., Highsmith, W.E., 1996. Heteroduplex generator in analysis of Rh blood group alleles. Clin. Chem. 42, 356−360.

Suzuki, Y., Orita, M., Shiraishi, M., Hayashi, K., Sekiya, T., 1990. Detection of ras gene mutations in human lung cancers by single-strand conformation polymorphism analysis of polymerase chain reaction products. Oncogene 5, 1037−1043.

Tannous, B.A., Verhaegen, M., Christopoulos, T.K., Kourakli, A., 2003. Combined flash- and glow-type chemiluminescent reactions for high-throughput genotyping of biallelic polymorphisms. Anal. Biochem. 320, 266−272.

Tiecke, F., Katzke, S., Booms, P., Robinson, P.N., Neumann, L., Godfrey, M., Mathews, K.R., Scheuner, M., Hinkel, G.K., Brenner, R.E., Hövels-Gürich, H.H., Hagemeier, C., Fuchs, J., Skovby, F., Rosenberg, T., 2001. Classic, atypically severe and neonatal Marfan syndrome: twelve mutations and genotype-phenotype correlations in FBN1 exons 24−40. Eur. J. Hum. Genet. 9 (1), 13−21.

Tominaga, T., 2007. Rapid determination of multi-locus sequence types of *Listeria monocytogenes* by microtemperature-gradient gel electrophoresis. J. Microbiol. Methods 70, 471−478.

Top, B., 1992. A simple method to attach a universal 50-bp GC-clamp to PCR fragments used for mutation analysis by DGGE. PCR Methods Appl. 2, 83−85.

Tully, L.A., Parsons, T.J., Steighner, R.J., Holland, M.M., Marino, M.A., Prenger, V.L., 2000. A sensitive denaturing gradient-gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region 1 of the human mtDNA control region. Am. J. Hum. Genet. 67, 432−443.

Uhrberg, M., Hinney, A., Enczmann, J., Wernet, P., 1994. Analysis of the HLA-DR gene locus by temperature gradient gel electrophoresis and its application for the rapid selection of unrelated bone marrow donors. Electrophoresis 15, 1044−1050.

van Elsas, J.D., Garbeva, P., Salles, J., 2002. Effects of agronomical measures on the microbial diversity of soils as related to the suppression of soil-borne plant pathogens. Biodegradation 13, 29−40.

Wang, Y.H., Barker, P., Griffith, J., 1992. Visualization of diagnostic heteroduplex DNAs from cystic fibrosis deletion heterozygotes provides an estimate of the kinking of DNA by bulged bases. J. Biol. Chem. 267, 4911−4915.

Watanabe, T., Saito, A., Takeuchi, Y., Naimuddin, M., Nishigaki, K., 2002. A database for the provisional identification of species using only genotypes: web-based genome profiling. Genome Biol. 3 (2). RESEARCH0010.

Waye, J.S., Nakamura, L.M., Eng, B., Hunnisett, L., Chitayat, D., Costa, T., Nowaczyk, M.J., 2002. Smith-Lemli-Opitz syndrome: carrier frequency and spectrum of DHCR7 mutations in Canada. J. Med. Genet. 39, e31.

White, M.B., Carvalho, M., Derse, D., O'Brien, S.J., Dean, M., 1992. Detecting single base substitutions as heteroduplex polymorphisms. Genomics 12, 301−306.

Wiese, U., Wulfert, M., Prusiner, S.B., Riesner, D., 1995. Scanning for mutations in the human prion protein open reading frame by temporal temperature gradient gel electrophoresis. Electrophoresis 16, 1851−1860.

Wood, N., Tyfield, L., Bidwell, J., 1993a. Rapid classification of phenylketonuria genotypes by analysis of heteroduplexes generated by PCR-amplifiable synthetic DNA. Hum. Mutat. 2, 131−137.

Wood, N., Standen, G., Hows, J., Bradley, B., Bidwell, J., 1993b. Diagnosis of sickle-cell disease with a universal heteroduplex generator. Lancet 342, 1519−1520.

Wood, N., Standen, G.R., Murray, E.W., Lillicrap, D., Holmberg, L., Peake, I.R., Bidwell, J., 1995. Rapid genotype analysis in type 2B von

Willebrand's disease using a universal heteroduplex generator. Br. J. Haematol. 89, 152−156.

Wu, D.Y., Wallace, R.B., 1989. The ligation amplification reaction (LAR) − amplification of specific DNA sequences using sequential rounds of template-dependent ligation. Genomics 4, 560−569.

Wu, D.Y., Ugozzoli, L., Pal, B.K., Wallace, R.B., 1989. Allele-specific enzymatic amplification of beta-globin genomic DNA for diagnosis of sickle cell anemia. Proc. Natl. Acad. Sci. U.S.A. 86, 2757−2760.

Xiao, W., Oefner, P.J., 2001. Denaturing high-performance liquid chromatography: a review. Hum. Mutat. 17, 439−474.

Yap, E.P., McGee, J.O., 1992. Nonisotopic SSCP and competitive PCR for DNA quantification: p53 in breast cancer cells. Nucleic Acids Res. 20, 145.

Yoshino, K., Nishigaki, K., Husimi, Y., 1991. Temperature sweep gel electrophoresis: a simple method to detect point mutations. Nucleic Acids Res. 19, 3153.

Zhang, Y., Coyne, M.Y., Will, S.G., Levenson, C.H., Kawasaki, E.S., 1991. Single-base mutational analysis of cancer and genetic diseases using membrane-bound modified oligonucleotides. Nucleic Acids Res. 19, 3929−3933.

Zhang, H.D., Zhou, J., Xu, Z.R., Song, J., Dai, J., Fang, J., Fang, Z.L., 2007. DNA mutation detection with chip-based temperature gradient capillary electrophoresis using a slantwise radiative heating system. Lab. Chip 9, 1162−1170.

Zhong, X.B., Reynolds, R., Kidd, J.R., Kidd, K.K., Jenison, R., Marlar, R.A., Ward, D.C., 2003. Single-nucleotide polymorphism genotyping on optical thin-film biosensor chips. Proc. Natl. Acad. Sci. U.S.A 100, 11559−11564.

Chapter 4

Quantitative Polymerase Chain Reaction

L. Overbergh, S. Vig, F. Coun and C. Mathieu
KU Leuven, Leuven, Belgium

4.1 HISTORY OF THE POLYMERASE CHAIN REACTION

The polymerase chain reaction (PCR) is a technique based on the exponential amplification of nucleic acids by the thermostable *Thermus aquaticus* (*Taq*) polymerase. The method uses a pair of synthetic oligonucleotide primers, each hybridizing to one strand of a double-stranded DNA (dsDNA) target of interest, with the pair spanning a region that will be exponentially amplified. The annealed primers act as a substrate for the *Taq* DNA polymerase, creating a complementary DNA strand via the sequential addition of deoxynucleotides (dNTPs), which have been added to the reaction mixture (Fig. 4.1). Experimentally, the process most typically consists of three steps:

1. a denaturation step at 94 or 95°C,
2. primer annealing to the single stranded DNA (ssDNA) strands at 55–65°C, and
3. a primer extension at 72°C.

By repeating these steps for a number of times, usually 30 to 40 cycles, the resulting DNA target sequence will be amplified exponentially, resulting in billions of copies of the so-called "amplicon." Based on this simple principle, a classical PCR amplification reaction will consist of three phases in time (Fig. 4.2):

1. Exponential amplification: At every cycle, the amount of product is doubled (assuming 100% reaction efficiency). The reaction is very sensitive and specific.
2. Linear phase (leveling off): The reaction slows down, due to decreasing activity of the *Taq* DNA polymerase and the consumption of reagents such as dNTPs and primers.
3. Plateau phase: The reaction has ended, and no more nucleic acids are synthesized due to the exhaustion of reagents and the *Taq* DNA polymerase.

The technique was first described by Kary Mullis in the 1980s (Faloona et al., 1986), for which he received the Nobel Prize in 1993. This classic end-point PCR is a yes/no reaction because it measures DNA product formation after a fixed number of cycles, that is, in the plateau phase of the reaction, thus resulting in qualitative information on the presence or absence of a certain gene or mRNA, but not revealing any information on the quantity of DNA or RNA. This PCR technique has become one of the most influential tools in the biological and medical sciences (Guyer and Koshland, 1989).

Over the years, numerous adaptations and applications to this classic end-point PCR have been described, including semiquantitative PCR and quantitative competitive PCR, before the introduction of real-time PCR. A first adaptation measured PCR product accumulation during the exponential phase of the reaction, resulting in semiquantitative data. This method necessitates interruption of the PCR reaction after an experimentally determined number of cycles. Furthermore, samples from a single experimental setup can be analyzed only over a relatively small linear range. Alternatively, competitive PCR has been developed, resulting in quantitative data. This method, however, needs extensive optimization, since it requires the coamplification of an internal cDNA or RNA control; the latter are specifically used in applications of reverse-transcriptase-quantitative PCR (qPCR; see later) (competitor) with the unknown sample in the same tube. In this case, quantification is performed by titrating an unknown amount of target template against a dilution series of known amounts of the standard. The internal control consists of target DNA or RNA that has been slightly modified. Thus one set of primers is designed that coamplifies the target and the competitor, with the same efficiency, although they can be distinguished from each other by, for instance, a difference in length or restriction sites

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00004-3
Copyright © 2017 Elsevier Ltd. All rights reserved.

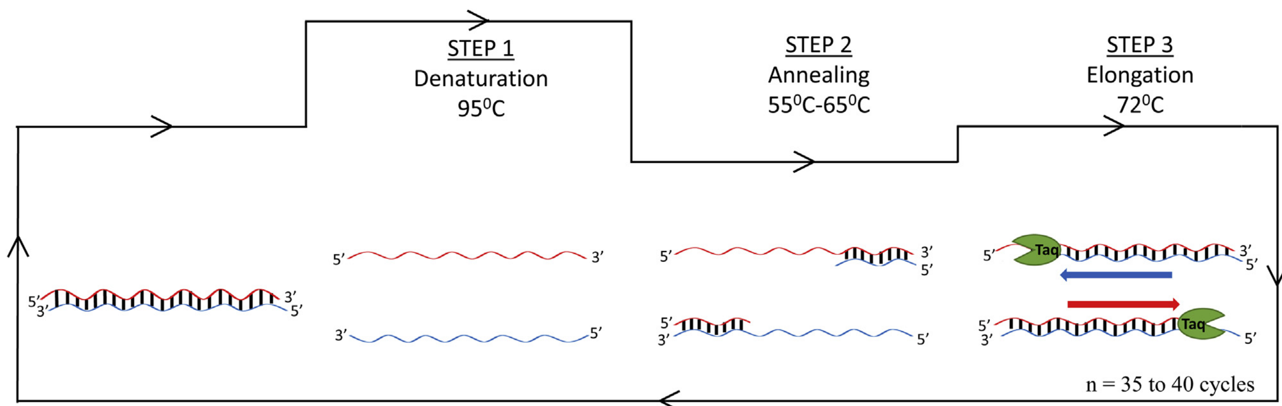


FIGURE 4.1 Principle of polymerase chain reaction (PCR). PCR is a technique based on the exponential amplification of nucleic acids. The method uses a pair of synthetic oligonucleotide primers, each hybridizing to one strand of a dsDNA target of interest, with the pair spanning a region that will be exponentially amplified by a *Taq* DNA polymerase, creating a complementary DNA strand. The process typically consists of three steps: (1) denaturation at 94 or 95°C; (2) primer annealing at 55−65°C; and (3) primer extension at 72°C. By repeating these steps for a number of times, usually 30 to 40 cycles, the resulting DNA target sequence will be amplified exponentially, resulting in billions of copies of the amplicon.

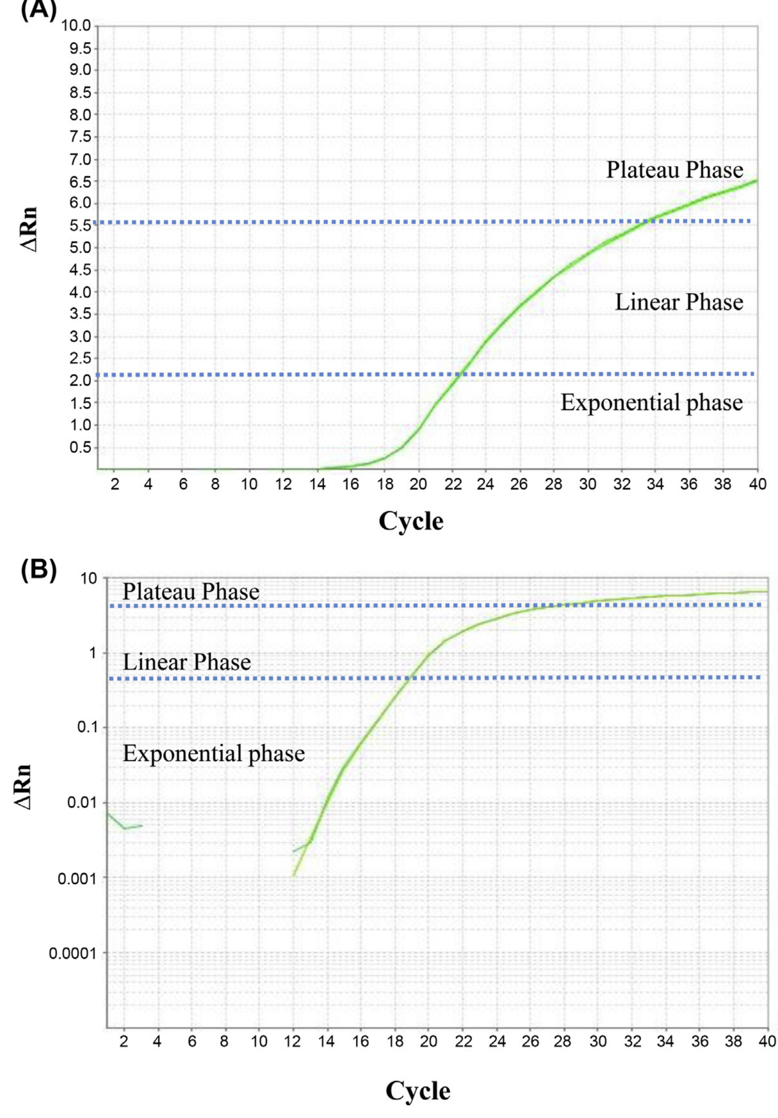


FIGURE 4.2 Typical course of a polymerase chain reaction (PCR) amplification plot. A PCR reaction consists of three phases: the exponential phase, the linear phase, and the plateau phase. In a classical PCR amplification plot, the number of cycles is plotted against the ΔRn signal (the difference between the fluorescence detected at a certain point of the reaction and the initial fluorescence or the fluorescent emission of the baseline), shown in the linear phase (A) or in the logarithmic phase (B). The threshold is chosen based on the variation of the baseline. The threshold cycle is the point where the detected fluorescence crosses this given threshold.

(Clementi et al., 1994). This method provides a strategy for accurate quantification, but the construction of internal standards is technically sophisticated and labor intensive.

For the detection of PCR products using either of these methods several detection techniques can be used, which all require excessive post-PCR manipulations. The most classically used are agarose gel electrophoresis with ethidium bromide or SYBR Green staining, fluorescent labeling and analysis using polyacrylamide gels, radioactive labeling, and Southern blotting or detection by phosphorimaging. Major drawbacks using these classical detection systems are the use of hazardous chemicals and the potential risk for laboratory contamination. Moreover, all of these post-PCR manipulations are very time-consuming.

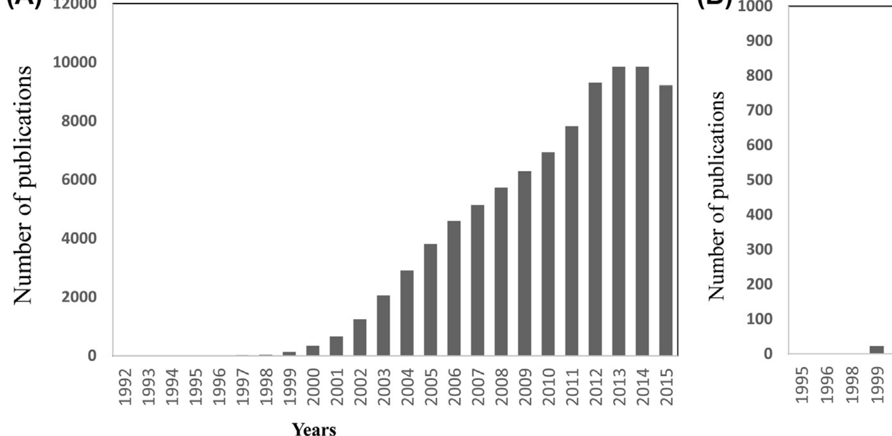
The development of a new procedure in the early 1990s for the analysis and quantification of DNA or RNA, based on fluorescence-kinetic real-time PCR, enabled the quantification of the PCR product in real-time (Higuchi et al., 1993; Gibson et al., 1996; Heid et al., 1996). This sensitive and accurate technique permits the quantification of the PCR product during the exponential phase of the PCR reaction. This is in full contrast to the classic end-point assays, as they are designed to provide information as rapidly as the amplification process itself, thus requiring no post-PCR manipulations. The development of this real-time (nowadays more often referred to as qPCR) technique again had a revolutionary impact on molecular biology. Indeed, the technique is widely used in research and in diagnostics, with countless researchers contributing to the exponential growth of PCR usage in a wide range of applications, such as DNA-, RNA- (coding mRNA, non-coding snRNA, and microRNA), or, more recently, protein- (immuno- or proximity ligation assay qPCR) based applications. Today, qPCR is recognized as the gold standard in the quantitative analysis of nucleic acids because of its high sensitivity and reproducibility, broad dynamic range, ease of use, and relatively low cost. Reverse transcriptase qPCR (RT-qPCR), another technique that allows quantitative measurements of RNA, was introduced at approximately the same time as qPCR. The latter consists of an additional step, namely, the conversion of RNA into complementary DNA by an enzymatic reverse transcriptase reaction. The RT-qPCR technique also follows the same exponential growth over time, although at a lower level (Fig. 4.3A and B).

4.2 PRINCIPLE OF REAL-TIME POLYMERASE CHAIN REACTION

About 10 years after the introduction of the classical end-point PCR technique, the introduction of dual-labeled oligonucleotide probes led to the invention of qPCR. Over the years, this new revolutionary technique has introduced different kinds of detection methods, which are all used today: hydrolysis probes, intercalating dyes, hybridization probes, prime–probe combinations, and modified nucleic acids.

4.2.1 Quantitative Polymerase Chain Reaction Using Hydrolysis Probes

Real-time PCR was first described using hydrolysis probes (Heid et al., 1996; Gibson et al., 1996), often referred to as the TaqMan system. The technique is based on the coupling of two important processes: first, the construction of dual-labeled oligonucleotide probes, also called hydrolysis or TaqMan® probes, which emit a fluorescent signal upon cleavage, based on the principle of fluorescence resonance energy transfer (FRET), a mechanism allowing energy transfer between two light-sensitive molecules through nonradiative dipole–dipole coupling (Stryer, 1978; Cardullo et al., 1988); and second, the discovery that the



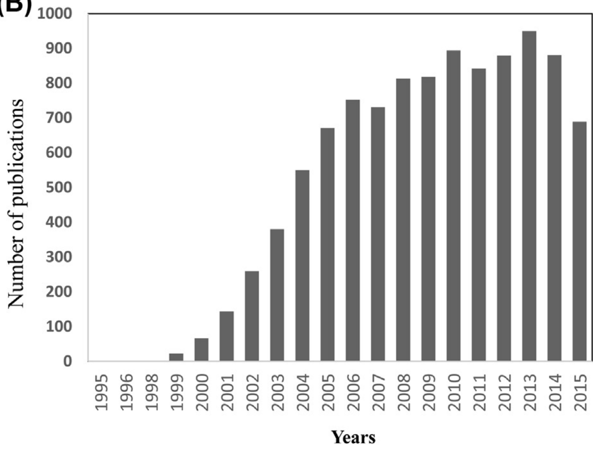


FIGURE 4.3 Real-time polymerase chain reaction (PCR) publications. The keywords "real time," "realtime," or "real-time" and "PCR"' (A) or "real time," "realtime," or "real-time" and "reverse transcription PCR" were typed into the PubMed search engine (http://www.ncbi.nlm.nih.gov/pubmed). The number of citations by year is shown. (A) qPCR publications; (B) RT-qPCR publicatons.

Taq DNA polymerase possesses, next to its 3′-> 5′ polymerase activity, a 5′-> 3′ exonuclease activity, which can be exploited to degrade the fluorescent labeled probe (Holland et al., 1991). The oligonucleotide probe used in this assay is nonextendable at its 3′ end and is dual-labeled with a reporter fluorochrome, for example, FAM (6-carboxyfluorescein), and a quencher fluorochrome, for example, TAMRA (6-carboxy-tetramethylrhodamine). It is designed to anneal to the target sequence internally of the primers during the annealing and extension phase of the PCR reaction. In its free, intact form, no fluorescent emission can be measured, because fluorescent emission of the reporter dye is absorbed by the quenching dye. However, upon annealing of the probe to one of the target strands, the probe will become degraded by the 5′-> 3′ exonuclease activity of the *Taq* polymerase. Consequently, the reporter and quencher dye become separated, and the reporter dye emission is no longer transferred to the quenching dye (no more FRET), resulting in an increase of reporter fluorescent emission (e.g., for FAM at 518 nm). This process occurs in every cycle and does not interfere with the exponential accumulation of the PCR product. The increase in fluorescence is measured cycle by cycle and directly correlates with the amount of PCR product formed (Gibson et al., 1996; Heid et al., 1996) (Fig. 4.4).

Other reporter dyes are available apart from the classically used fluorescent reporter dye FAM. These include TET (tetrachloro-6-carboxyfluorescein), JOE (2,7,-dimethoxy-4,5-dichloro-6-carboxyfluorescein), HEX (hexacholoro-6-carboxyfluorescein), VIC, Texas Red, or Cy5. The choice of different reporter dyes, with a minor overlap in the fluorescent emission spectra, makes it possible to perform multiplex PCR reactions, thus simultaneously amplifying different DNA targets. Similarly, there is a choice between different quencher dyes. The most classically used quencher dye is TAMRA. DABCYL (4-(4Á-dimethylaminophenylazol) benzoic acid) also can be used as a quencher dye, but its use is much more prevalent in the molecular beacons probes (see later). An advantage of using DABCYL is its reduced autofluorescence compared to TAMRA. In addition, several dark quencher fluorochromes have become available, which are used more frequently than the classical quenchers. They absorb the energy that is emitted by the reporter dye and release it as heat rather than as fluorescence. This results in a lower background signal and thus a higher sensitivity.

Another probe design of special interest are minor groove binding probes (MGB). These are, like the TaqMan probes, hydrolysis probes with a minor groove binding molecule attached to the end of the probe, resulting in a higher affinity to the complementary DNA target region, allowing for the construction of shorter probe designs with increased specificity. These are more sensitive, especially to single base mismatches, and are therefore ideally suitable for single nucleotide polymorphism (SNP) detection and allelic discrimination (Kutyavin et al., 2000).

4.2.2 Quantitative Polymerase Chain Reaction Using DNA Intercalating Dyes

A second widely used qPCR technique is based on the detection and quantification of PCR products using fluorescent DNA intercalating dyes. The principle of this technique was first described by Higuchi (Higuchi et al., 1993), who monitored the increase in ethidium bromide fluorescence using a charge-coupled device camera, a method that was referred to as kinetic PCR. Shortly thereafter, SYBR Green I (N′,N′-dimethyl-N-[4-[(E)-(3-methyl-1,3-benzothiazol-2-ylidene)methyl]-1-phenylquinolin-1-ium-2-yl]-N-propylpropane-1,3-diamine), an asymmetric cyanine dye that is less toxic than ethidium bromide, became widely used as a dye that incorporates into dsDNA (Zipper et al., 2004). In the meantime, a whole range of other asymmetric cyanine dyes became available, such as SYBR Green Save, EvaGreen, SYBR Gold, SYTO, BEBO (4-[(3-methyl-6-(benzothiazol-2-yl)-2,3-dihydro-(benzo-1,3-thiazole)-2-methylidene)]-1-methyl-pyridinium iodide) (Bengtsson et al., 2003), and BOXTO (4-[6-(benzoxazole-2-yl-(3-methyl-)-2,3-dihydro-(benzo-1,3-thiazole)-2-methylidene)]-1-methyl-quinolinium chloride) (Ahmad,

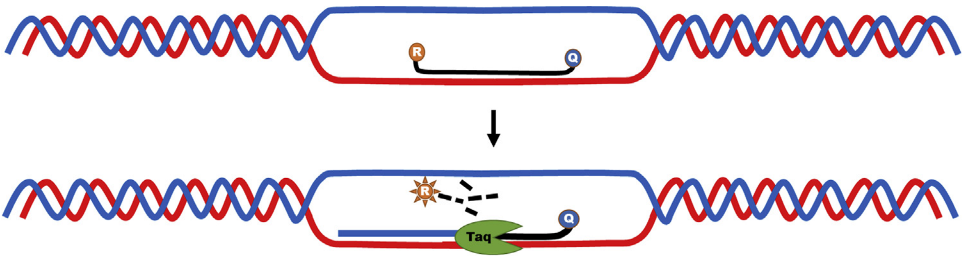


FIGURE 4.4 Hydrolysis or TaqMan probes. The dual-labeled TaqMan probe is cleaved by the 5′ exonuclease activity of the *Taq* DNA polymerase during the extension step of the polymerase chain reaction. The reporter fluorophore (R) and quencher (Q) are separated, which results in an increase in fluorescent emission.

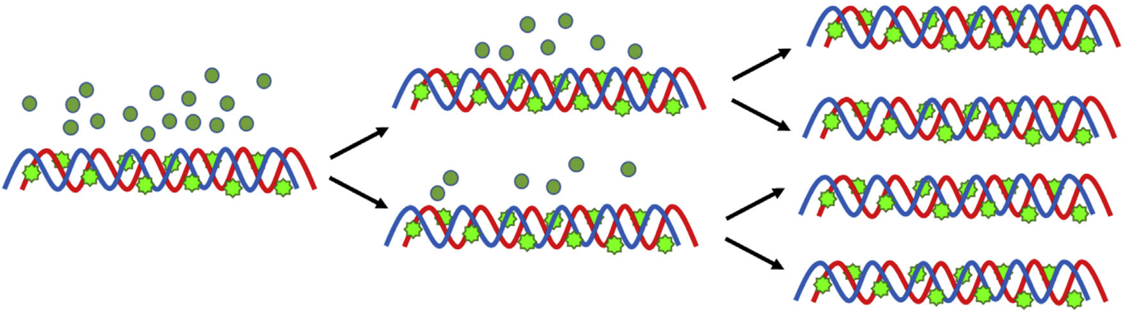


FIGURE 4.5 DNA intercalating dyes. DNA intercalating dyes incorporate into the minor groove of dsDNA. In solution no fluorescent light is emitted (*green circle*), but when incorporated into dsDNA the dye will emit fluorescent light (*green star*).

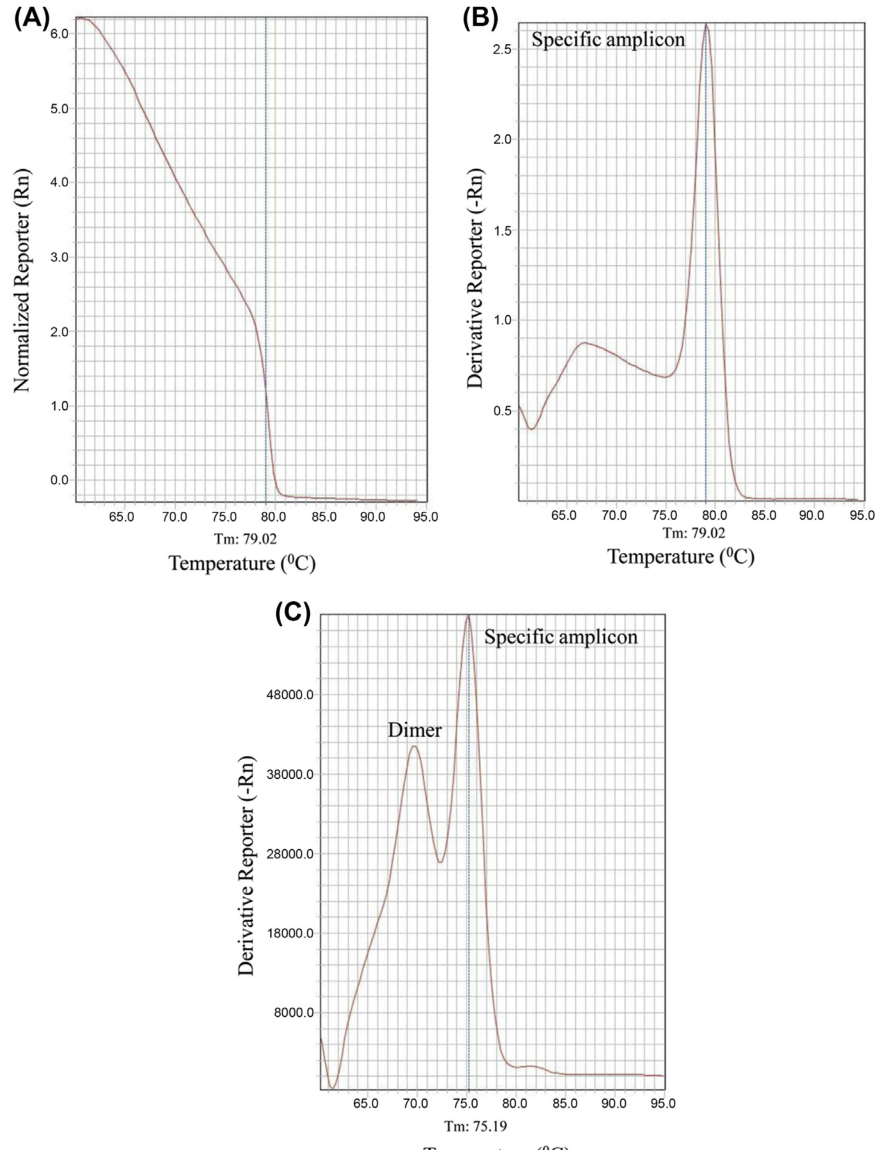


FIGURE 4.6 Post-polymerase chain reaction (PCR) melting curve analysis when using DNA intercalating dyes. By slowly increasing the temperature after the PCR amplification reaction from 60°C to 95°C, the dye fluorescence drops rapidly when the dsDNA melts into ssDNA strains (A). The first negative derivative of the fluorescence (y-axis) is plotted against the temperature (x-axis) (B and C). In the case of specific DNA amplification a single peak (peak one) is observed (B), while in the case of primer-dimer formation a second peak at a lower temperature (peak two) is observed (C). *Tm*, melting temperature.

2007). All of these dyes incorporate into the minor groove of dsDNA, by which its fluorescence is greatly enhanced. During the PCR reaction, the amount of double-stranded target will increase exponentially, paralleled by an increase in dye incorporation and fluorescent emission. In each cycle the fluorescent emission will increase gradually during the extension phase of the reaction and will be low or absent during the denaturation phase (Fig. 4.5). Different disadvantages have been reported for the use of SYBR Green I, as opposed to other DNA intercalating dyes, such as inhibition of PCR amplification in a concentration-dependent manner, effects on the DNA melting temperature, and preferential binding to specific DNA sequences (Gudnason et al., 2007). Despite this, SYBR Green I still remains the most widely used intercalating dye.

The greatest advantage of the use of intercalating dyes as opposed to the use of fluorescent labeled probes or primers is that they can be used with any pair of primers for any target. Because of this, the method can be easily used for multiplex PCR assays. Moreover, it is a cheaper alternative and requires less specialist knowledge than the design of fluorescent labeled probes. Consequently, the specificity is diminished due to the risk of amplifying nonspecific PCR products or primer dimers. After termination of the PCR reaction a melting curve analysis has to be performed to discriminate between specific and nonspecific PCR products (Ririe et al., 1997). In this way, the fraction of fluorescence originating from the specific target can be distinguished from that originating from primer dimers or nonspecific amplification products. This analysis is performed by slowly increasing the temperature from 60°C to 95°C, during which fluorescent emission is monitored continuously. Fluorescent emission will be high at low temperatures (when all PCR products are double stranded), and it will decrease dramatically around the melting temperature of the PCR product. The rationale behind this melting curve analysis is based on the fact that PCR products of different lengths and/or guanine-cytosine content will have different melting temperatures, which will result in distinct peaks when plotting the first negative derivative of the fluorescence versus the temperature. Because of their shorter length, primer dimers will, for instance, have a lower melting temperature as compared to the specific PCR product. This will consequently lead to a fluorescent peak at a lower temperature when plotting the first negative derivative of the fluorescence versus the temperature. On the other hand, a PCR product resulting from, for instance, contaminating genomic DNA, will have a higher melting temperature, which will again be reflected in the first negative derivative of the fluorescence versus the temperature (Fig. 4.6).

4.2.3 Quantitative Polymerase Chain Reaction Using Hybridization Probes

Quantitative PCR analysis with hybridization probes uses two juxtaposed sequence-specific probes, also known as HybProbes. The development of this system was first described in the mid-1980s (Heller and Morrison, 1985). Each probe has a single fluorescent reporter; the first one is a donor fluorophore at its 3′ end, and the other is an acceptor fluorophore at its 5′ end. The sequences of the two probes are designed to anneal to the target sequences in very close proximity to each other (i.e., within 1−5 nucleotides), in a head-to-tail arrangement, bringing the two dyes very close to each other. As long as the probe is in its free, unbound form, the fluorescent signal from the reporter dye is not detected. However, during the annealing phase of the PCR reaction, the probes anneal to the target sequence, and the two fluorophores come in close proximity to each other. This will result in the emission of light from the donor fluorochrome, which will excite the acceptor fluorochrome, a process referred to as resonance energy transfer. The dye in one of the probes transfers energy, allowing the other one to dissipate fluorescence at a different wavelength. The amount of fluorescence emitted can be measured during the annealing phase of the PCR reaction and is directly proportional to the amount of target DNA generated during the PCR process (Bernard and Wittwer, 2000) (Fig. 4.7). Hybridization probes are usually constructed with FAM as the 3′ donor fluorophore, and a range of different acceptor fluorophores are commonly used (e.g., ROX, Cy5, LC Red640, LC Red705) as the 5′ acceptor fluorophore.

Another form of hybridization probes are the molecular beacons, which contain a stem-and-loop structure in their intact, unbound form. They are dual-labeled, with a fluorophore linked to one end of the molecule and a quencher linked to the other end (Tyagi and Kramer, 1996). Fluorescence is quenched when the probe is in its hairpin-like structure due to the proximity between the quencher and the fluorophore, resulting in the complete absorption of any photons emitted by the fluorophore. When the probe sequence in the loop anneals to a complementary target sequence, a conformational change allows the formation of a linear structure whereby fluorescent energy transfer no longer occurs, resulting in an increase in fluorescence emission (Fig. 4.8). Molecular beacons are especially suitable for identifying point mutations. They can distinguish targets that differ by only a single nucleotide, and they are significantly more specific than conventional hydrolysis probes of equivalent length.

Other examples of hybridization probes that have been developed are HyBeacon™, MGB-Pleiades, MGB-Eclipse, ResonSense®, Yin-Yang, or displacement probes.

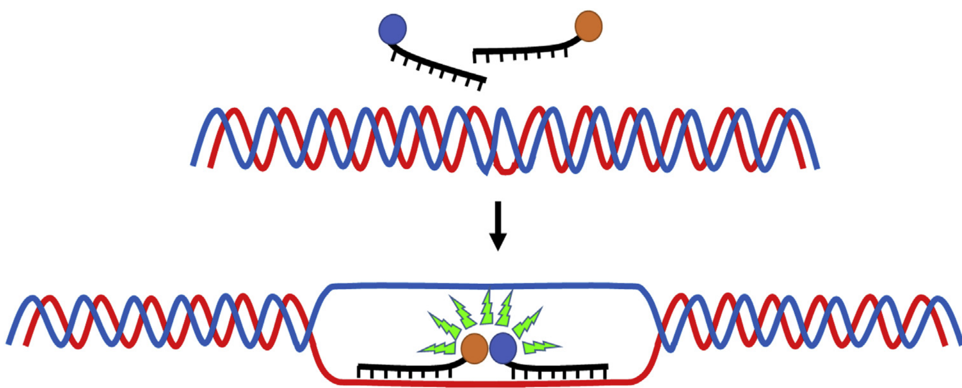


FIGURE 4.7 **Hybridization probes.** Two probes are used: one carrying an acceptor fluorophore, and the other carrying a donor fluorophore. When both fluorophores are brought into close proximity, that is, when the probes anneal to the target sequence, the donor is able to excite the acceptor through fluorescence resonance energy transfer, and fluorescence emission will occur.

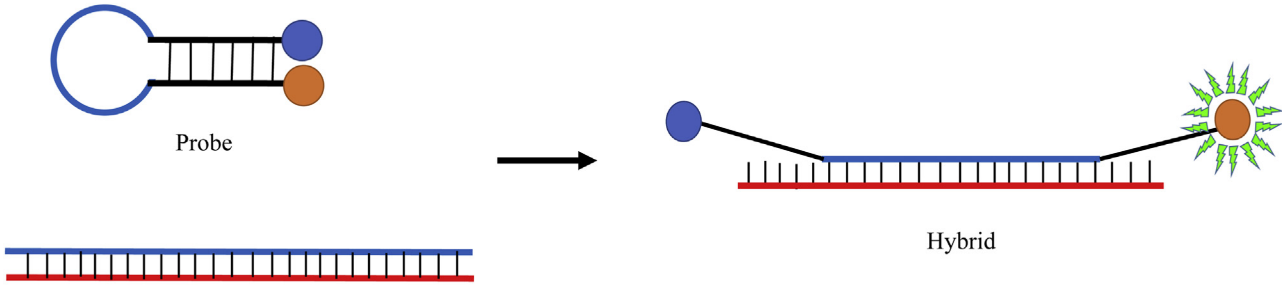


FIGURE 4.8 **Molecular beacons.** Molecular beacons are hairpin-shaped probes. The fluorophore and quencher are in close proximity when the probe is in its free, unbound state. When the probe anneals to the complementary target sequence, its conformation will change. The quencher and fluorophore are separated, causing fluorescence emission.

4.2.4 Quantitative Polymerase Chain Reaction Using Primer Probes

A variant of the classic hybridization probes uses a fluorescent labeled primer/probe combination. In this case, a 5′-labeled hybridization probe is designed to anneal to the PCR strand in close proximity to one of the PCR primers, which has a fluorophore at its 3′ end. This method requires that the fluorescently labeled primer be positioned near the probe, usually within five base pairs, to allow for adequate resonance energy transfer with the complementary probe (Von Ahsen et al., 2000). In this category different types of primer-probe systems have been developed, such as the Scorpions®, Amplifluor® probes or Sunrise system, LUX™, Cyclicons, ResonSense®, and Angler® probes (Lee et al., 2002), all making use of two oligonucleotides: a primer and a fluorescent molecule, which combines the primer and probe functions. The two oldest developments in this group are the Scorpion® system (Whitcombe et al., 1999) and the Amplifluor® probes (Nazarenko et al., 1997). In these primer/probe combinations, a primer sequence is linked to a specific probe sequence that is held in a hairpin-loop form. The stem-loop tail is separated from the primer by a PCR blocker to prevent the *Taq* DNA polymerase from amplifying the stem-loop sequence. This configuration brings the fluorophore into close proximity to the quencher and avoids fluorescence. As soon as annealing between the primer/probe and the target occurs, the hairpin is opened, and the fluorophore and quencher are separated, resulting in an increase in fluorescence emission (Fig. 4.9). In the case of LUX™ probes, which also have a hairpin structure, the 3′-end acts as a primer, which contains a single reporter fluorescent dye (Nazarenko et al., 2002) but does not require the presence of an internal quencher. Scorpions® differ from molecular beacons and hydrolysis probes in that their structure promotes a unimolecular probing mechanism. This results in a stronger fluorescent signal, especially under fast cycling conditions (Thelwell et al., 2000).

4.2.5 Real-Time Polymerase Chain Reaction Using Modified Nucleic Acids

In all of the different assay designs described earlier, the labeled primers or probes are synthesized based on

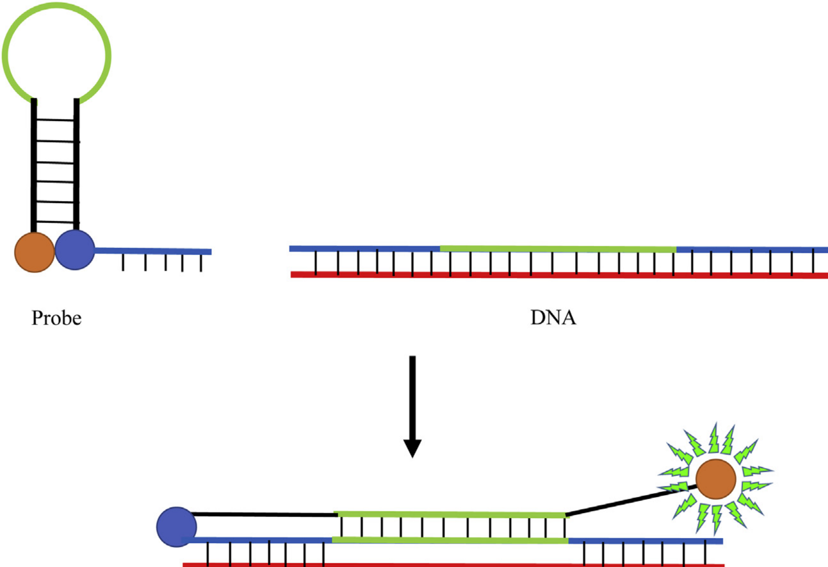


FIGURE 4.9 Scorpions. Scorpions are single-stranded, dual-labeled fluorescent primer/probes with a hairpin-shaped structure. The primer/probe contains a 5′ end fluorophore and an internal quencher dye directly linked to the 5′ end of a polymerase chain reaction (PCR) primer via a PCR blocker. In the unbound form, no fluorescent emission occurs. When binding to the target sequence, the hairpin shape unfolds, and the loop region of the probe hybridizes intramolecularly to the newly synthesized target sequence. This results in fluorescent emission, because the fluorophore and quencher are separated.

classical nucleic acids. For specific applications synthetic analogues are also used. This is the case in the locked nucleic acid (LNA) probes. LNA is a chemically modified RNA molecule in which the ribose moiety has an extra bridge connecting the 2′ oxygen and the 4′ carbon (Fig. 4.10A). Such modification significantly increases the hybridization affinity of the probe with DNA or RNA. Due to this capacity, very short probes can be designed, resulting in a very high binding efficiency to their complementary DNA, making them particularly suitable for SNP detection or miRNA quantification (Bonetta, 2005). An example of LNA probes are the tiny molecular beacons, which are shorter probes than the classical molecular beacons described earlier (Catrina et al., 2012). More recent developments making use of chemically modified nucleic acids include the peptide nucleic acids (PNAs). PNAs are artificially synthesized polymers composed of a backbone of repeating N-(2-aminoethyl)-glycine units linked by peptide bonds (Nielsen et al., 1991) (Fig. 4.10B). The probe is an uncharged molecule due to the absence of phosphate groups. Therefore it hybridizes faster and binds with a much higher affinity to DNA or RNA, forming stronger duplexes as compared to DNA/DNA duplexes. As for LNA probes, this makes

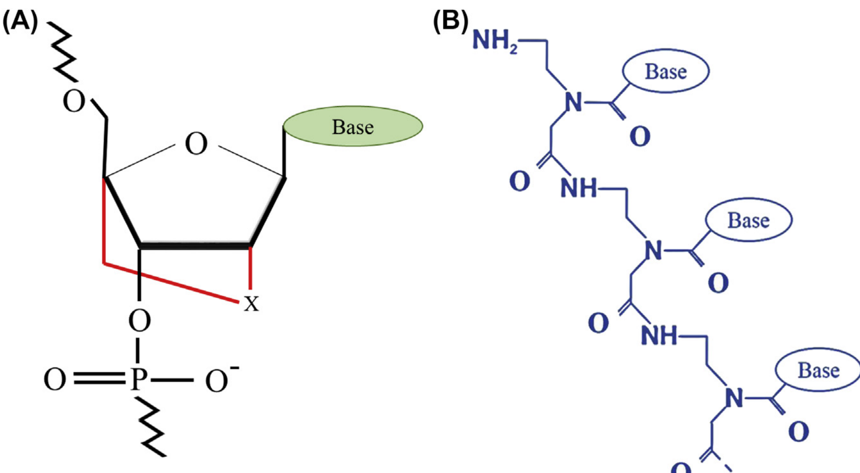


FIGURE 4.10 Structure of locked nucleic acids (LNAs) and peptide nucleic acids (PNAs). Chemical structure of LNA (A) and PNA probes (B).

shorter probe designs possible with a very high specificity. An example of a PNA probe is the light-up probe, a PNA probe to which an asymmetric cyanine dye is attached. Upon probe hybridization the dye binds to the target DNA, which results in a large enhancement in dye fluorescence (Svanvik et al., 2000).

Another example of modified oligonucleotides with improved hybridization properties are the Zip Nucleic Acids (ZNA™), which are designed by conjugating spermine derivatives as cationic units to an oligonucleotide. This results in a decrease of electrostatic repulsions due to the polyanionic nature of the nucleic acids, thereby increasing the affinity of the probe to its DNA target. Like LNA and PNA probes, they are very well suited for applications like SNP and miRNA detection and the detection of very low abundant targets (Moreau et al., 2009; Paris et al., 2010).

A totally different approach is used in the Plexor™ technology, which makes use of two modified bases, isoguanine and 5′-methylisocytosine. These bases are known to form a unique base pair interaction when incorporated into dsDNA, as they can only pair with each other. One PCR primer is synthesized with an iso-dC residue and a fluorescent label at the 5′ end. The inclusion of dabcyl-modified iso-dG nucleotides in the reaction mix will result in quenching of the fluorescent signal upon incorporation of this nucleotide at the position complementary to the iso-dC residue (Gašparič et al., 2008).

4.3 REAL-TIME THERMAL CYCLERS

Since the development of qPCR in the mid-1990s, instrumentation systems have undergone an extensive evolution. The first available instrument, the 7700 SDS from Applied Biosystems, took up almost an entire bench area, made use of an expensive laser as the light source, and had to be placed in an air-conditioned laboratory. This instrument was unable to produce a true analysis in real-time, since data could be viewed only after termination of the PCR reaction. As the technology improved, systems became available that allowed the detection of a PCR product from the moment it was formed, thus in pragmatic real-time. These newer machines are much smaller in size, and the laser has been changed for less precious tungsten-halogen or light-emitting diode lamps. Even portable backpack-sized instruments have been developed, which can be used for on-site analysis. These instruments are able to give results in less than 30 min of analysis time, a rather important feature, for instance, for the on-site detection of pathogen outbreaks. Because of the wide range of instruments developed by different companies, prices have dropped significantly, and instruments are becoming a standard tool for routine molecular and diagnostic laboratories.

Although all possible applications can be performed on all instruments, each one has specific advantages and disadvantages. The choice for a specific real-time thermal cycler is therefore dependent on the specific application one is focusing on. For the quantification of gene expression, for instance, the length of a run is not the most important factor, although one would like to choose an instrument that can analyze a large number of samples simultaneously. Indeed, for a high sample throughput, a 96-well or even a 384-well system will be the preferable choice. On the other hand, if the main application is, for example, pathogen identification, velocity is the main issue, and the best choice would be a rapid thermal cycler. Most of the thermal cyclers are able to perform multiplex PCR, are sensitive and accurate, and increasingly present user-friendly software. The rapid evolution in the development of these instruments indicates the competitiveness and widespread use of PCR technology.

4.4 HOW DATA ARE OBTAINED

When performing qPCR, the ability to monitor the amplification process of the PCR reaction in real-time revolutionizes the way in which the data are obtained. Typically, the reactions are characterized by the point in time during the exponential phase of PCR cycling when amplification of a PCR product reaches a certain detection level, as opposed to end-point detection, where the amount of product formed is measured after a fixed number of cycles, most often in the plateau phase of the reaction. Furthermore, quantification is based on the inherent property of a PCR reaction, that the more input DNA copies one starts with, the fewer cycles of PCR amplification it takes to make a specific number of amplification products. Finally, the fact that the formation of amplification products linearly correlates with the amount of fluorescence emission is exploited in the qPCR assay (Giulietti et al., 2001).

In practice, using any of the developed detection chemistries on any of the available instruments, the increase in fluorescence emission can be read by a sequence detector in real-time during the course of the reaction and is a direct consequence of target amplification during PCR. In Fig. 4.11, a typical amplification plot is shown in which the terms and definitions routinely used in qPCR are illustrated. During the initial cycles of the PCR reaction, there is little or no change in the fluorescence signal. This stage is the baseline of the amplification plot. The fluorescence emission of the product at each time point is measured during PCR cycling and is defined as Rn^+ (Fig. 4.11A). Analogously, Rn^- is the fluorescent emission of the baseline. The increase in fluorescence is calculated by the computer software program and is plotted on the y-axis as the ΔRn value, using the equation $\Delta Rn = Rn^+ - Rn^-$ (Fig. 4.11B). Thus this value directly correlates with probe degradation

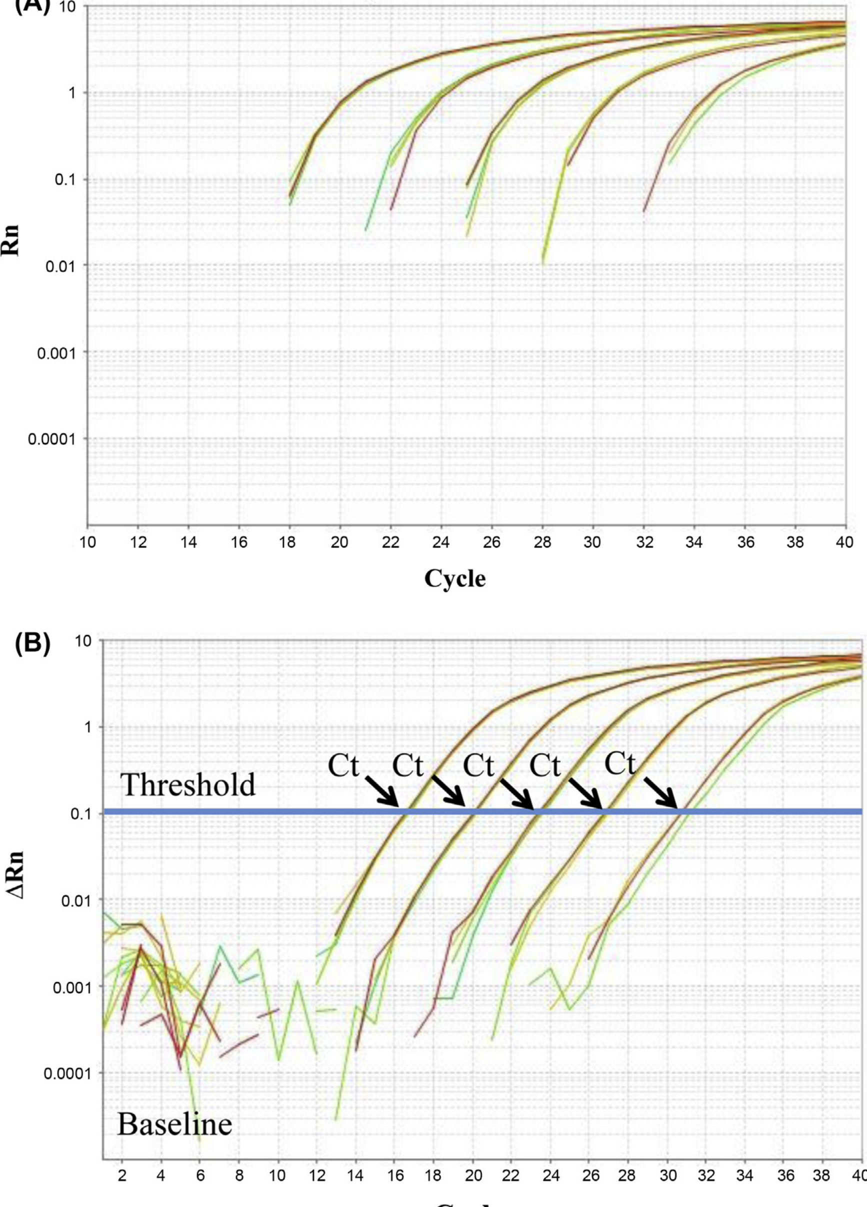


FIGURE 4.11 **Polymerase chain reaction (PCR) amplification plot of a dilution series.** Five points of serial 10-fold dilutions of input DNA amplified by quantitative polymerase chain reaction. (A) The fluorescence emission of the product at each time point is measured during PCR cycling and is defined as Rn^+. (B) The increase in fluorescence is plotted on the y-axis as the ΔRn value, using the equation $\Delta Rn = Rn^+ - Rn^-$. A threshold is defined (*blue line*). Ct = cycle threshold.

(in the case of hydrolysis or TaqMan probes) or with the incorporation of dyes (in the case of asymmetric fluorescent dyes) during the PCR process and consequently with the formation of a specific PCR product. An arbitrary threshold is chosen, based on the variability of the baseline, usually determined as 10 times the standard deviation of the baseline. This, of course, can be changed manually for each individual experiment if necessary. Threshold cycle (Ct) values are then calculated by determining the point at which the fluorescence exceeds this chosen threshold. Ct is reported as the cycle number at this point. Therefore Ct values decrease linearly with increasing input target quantity. This is used as a quantitative measurement of the input target.

4.5 HOW DATA ARE QUANTIFIED

Two different methods are commonly used to quantify the results obtained by qPCR: the standard curve method and the comparative threshold method. A third elegant method

is the Pfaffl method, although its use is less widely spread (Giulietti et al., 2001).

4.5.1 Standard Curve Method

In the standard curve method, a sample with a known concentration is used to make a dilution series, which is used as a standard curve. Samples, which can be used to construct a dilution series, are purified plasmid dsDNA, in vitro transcribed RNA, in vitro synthesized ssDNA, or any cDNA sample expressing the target gene. The concentration of these DNA or RNA samples can be measured spectrophotometrically at 260 nm and converted to the number of copies using the molecular weight of the DNA or RNA. For absolute quantification of mRNA expression absolute standards have to be used (for instance, in vitro transcribed RNA). Such absolute standards provide a very high accuracy and are often used in International Organization for Standardization-accredited laboratories (Saitta et al., 2016). Often, in this regard, a standard reference material is used in order to standardize assays between different laboratories (Hartnell et al., 2012). More often, that is, in nonaccredited laboratories, typically research laboratories, cDNA plasmid standards are used for quantification. These are constructed by cloning a cDNA fragment into a suitable plasmid vector. For the quantification of mRNA expression, however, this will result in only a relative quantification, because variations in efficiency of the reverse transcription step are not controlled.

When creating a standard curve by serial 2-fold dilutions of the standard sample, two consecutive points will have a Ct difference of 1. Similarly, Ct values from 10-fold diluted samples will differ by 3.3. This, of course, is assuming that 100% PCR efficiency is reached. Consequently, the slope of the standard curve is a measure of the efficiency of the PCR reaction. For serial 10-fold dilutions, it should ideally be −3.3. In practice, standard curves with a slope between −3.0 and −3.6 are considered acceptable. Also, the sensitivity of the PCR reaction is reflected in the standard curve by the point at which the standard curve crosses the y-axis (Y-intercept). Indeed, the lower the Ct value at this point, the higher the sensitivity of the PCR reaction (Fig. 4.12). By plotting the Ct value of an unknown sample on the standard curve, the amount of input target sequence in the sample can be determined. Usually, this calculation is performed automatically by the software program of the qPCR instrument.

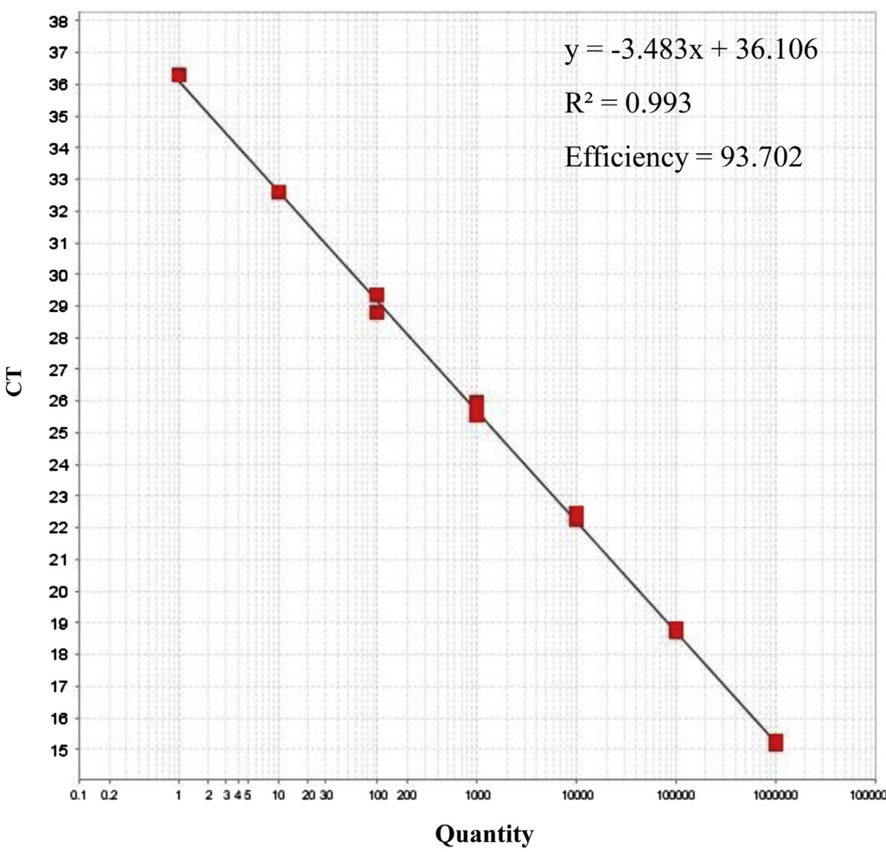


FIGURE 4.12 **Standard curve.** Seven points of a 10-fold dilution series of cDNA, performed in triplicate wells, amplified using the TaqMan system. A standard curve is generated by plotting threshold cycle (Ct) values against relative input cDNA copy numbers.

4.5.2 Comparative Threshold Cycle Method

An alternative method used for relative quantification is the comparative Ct method, or the ΔΔCt method (Livak and Schmittgen, 2001). This method uses arithmetic formulas to calculate relative expression levels, compared to a calibrator. A nontreated control sample can be used, for instance, as a calibrator. Moreover, the value of the unknown target is normalized to an endogenous reference gene (e.g., a housekeeping gene, discussed later). The amount of target, relative to the calibrator and normalized to the reference gene, is measured by the equation $2^{-\Delta\Delta Ct}$, where $\Delta\Delta Ct = \Delta Ct\text{sample} - \Delta Ct\text{calibrator}$, and ΔCt is the Ct of the target gene subtracted by the Ct of the reference gene. The equation thus represents the normalized expression of the target gene in the unknown sample, relative to the normalized expression of the calibrator sample. Importantly, for the ΔΔCt method to be applicable, the efficiency of PCR amplification for the target gene and the reference gene must be approximately equal. For every qPCR assay that is being set up this has to be tested by determining how the ΔCtsample and the ΔCtcalibrator vary with template dilution. In case the efficiencies between PCR amplification for the target and reference genes are different, a new set of primer/probe combinations has to be designed. Alternatively, the standard curve method or the more recently developed Pfaffl method can be applied.

4.5.3 Pfaffl Method

As an alternative to the comparative Ct method, another mathematical model has been developed by M.W. Pfaffl (Pfaffl, 2001). In this method, the PCR efficiency of the target and reference genes is taken into account. Relative expression ratios (R) are calculated based on the efficiency of the qPCR assay (E) and the threshold cycle of the unknown sample (Ct) versus a calibrator (e.g., a nontreated control sample) and expressed in comparison to a reference gene. This is expressed by the following equation:

$$\text{Ratio} = (1 + E_{\text{target}})^{\Delta Ct(\text{target})} \Big/ (1 + E_{\text{ref}})^{\Delta Ct(\text{ref})}$$

where E_{target} is the PCR efficiency of the target gene; E_{ref} is the PCR efficiency of the reference gene; $\Delta Ct(\text{target}) = Ct(\text{sample}) - Ct(\text{calibrator})$ of the target gene; and $\Delta Ct(\text{ref}) = Ct(\text{sample}) - Ct(\text{calibrator})$ of the reference gene.

In analogy with the ΔΔCt method, ΔCt is the Ct of the target gene subtracted by the Ct of the reference gene. Because this method takes into account the intrinsic efficiency of each PCR reaction, it can be applied for any qPCR amplification assay with enhanced accuracy. A major drawback is that in this method one cannot make use of the mean of different reference genes (see also later).

4.6 MULTIPLEX QUANTITATIVE POLYMERASE CHAIN REACTION

Multiplex qPCR can refer either to the simultaneous amplification and detection of different target genes in one tube or to the use of multiple fluorogenic probes for the discrimination of different alleles. The term is therefore somewhat confusing, since it is not a mere extension of classical multiplexing known in conventional PCR, which was simply the amplification of multiple templates within one reaction using different primers.

Since diagnostic analyses are often restricted by the limited availability of bioptic material, and one of the primary goals is to decrease analysis time, multiplex PCR was considered to be an attractive solution, especially in the field of infectious diseases. As such, major efforts have been undertaken in order to optimize multiplex diagnostic detection kits based on nucleic acid amplification. One obstacle that has to be dealt with arises from the limited number of different fluorophore reporters with a good spectral resolution. qPCR instruments contain optimized filters to minimize the overlap of the emission spectra from the different available fluorophores. Despite this, the number of fluorophores that can be combined and clearly distinguished is limited when compared with the resolution in conventional multiplex PCR. Improvements in the design of other probe formats as well as novel combinations of fluorophores have greatly aided the ability to simultaneously detect a larger number of targets. Also, developments in qPCR instruments have significantly improved, allowing for multiplexing of up to six different colors. Another approach, which is used for the detection of human genetic diseases, is based on the discrimination between single or multiple nucleotide changes (between different alleles) by making use of the differences in melting points. In this approach, a single fluoroprobe can be used to distinguish between products, based on their distinct melting temperatures, which is reflected in the differences in thermodynamic stabilities of the perfectly complementary and the mismatched probe-target duplexes. Finally, the combined use of multicolor fluorimetry and fluorescence melting curve analysis can greatly increase the number of targets that can be detected simultaneously (Wittwer et al., 2001).

Based on these separate or combined approaches, the use of multiplex PCR in diagnostic kits, approved by the Food and Drug Administration, has increased greatly over the last years and is becoming a standard method in many clinical settings. Major advances compared to classical culture and microscopic technologies are its enhanced sensitivity and specificity and reduced analysis time. However, one needs to keep in mind that this technology is limited to the existing knowledge of a microorganism genome, that no distinction can be made between life and

death organisms, and that unexpected mutations, which often take place in microorganisms, may not be detected. As such, a combination of classical diagnostic techniques combined with multiplex PCR will probably remain as the standard in most clinical diagnostic settings (Gray and Coupland, 2014).

4.7 APPLICATIONS OF QUANTITATIVE POLYMERASE CHAIN REACTION AND REVERSE TRANSCRIPTASE-QUANTITATIVE POLYMERASE CHAIN REACTION

The qPCR method is widely used in fundamental research laboratories as well as in clinical diagnostic settings with a variety of applications either based on the amplification of DNA (qPCR) or RNA (RT-qPCR). The latter refers to the combined reverse transcriptase (RT) and PCR amplification, used for either the amplification of coding or non-coding RNA.

4.7.1 Clinical Microbiology

There is a growing interest in the use of real-time PCR for the detection of pathogens for the diagnosis of viral, bacterial, or fungal infections (De Crom et al., 2016; Stappers et al., 2016). Most of the assays developed allow for an increased frequency as well as enhanced speed of pathogen detection as compared to conventional culture techniques. Moreover, the quantification of pathogen load is possible. For real-time assay development, one has to take into account that most infectious agents are characterized by a high mutation rate, which can dramatically influence the pathogen load estimation. This can be overcome by designing primers in highly conserved regions. Alternatively, of course, sequence variations can provide the basis for the development of subtype-specific assays. It remains of crucial importance to further improve real-time PCR-based assays in terms of their sensitivity, the efficiency of DNA extraction, and especially the capacity to detect a much broader range of pathogens using this technology (Warhurst et al., 2015). Also, the rapid evolution of resistance in Gram-negative bacteria necessitates new methods that allow for rapid detection (Endimiani and Jacobs, 2016).

4.7.2 Food and Industrial Microbiology

The rapid and accurate detection of bacterial or viral pathogens in food samples is important both for food quality assurance and to trace outbreaks within food supplies. Real-time PCR assays are being optimized more often for the detection of these pathogens, since they significantly shorten the analysis time compared to conventional biochemical and serological identification methods and because the technique can be applied directly on preenrichment media or food products. A specific problem when analyzing food pathogens is that foods are often complex matrices, which require selective enrichment steps to overcome problems of low pathogen numbers. To address this issue, a universal enrichment broth has been developed, enabling the enrichment of multiple pathogens (Bailey and Cox, 1992). An additional problem that has to be faced in the case of PCR-based detection methods for food pathogens is to discriminate between life and death pathogenic cells. One approach is to apply a living/dead dye that covalently binds to DNA and inhibits PCR amplification from dead cells (Rudi et al., 2002). Alternatively, the development of mRNA-based real-time PCR assays (as opposed to DNA-based assays) can be applied, which has the advantage of serving as an accurate indicator of pathogen viability (Rijpens and Herman, 2002). Protocols have been validated for the detection of *Escherichia coli*, *Salmonella* (Bhagwat, 2003), *Listeria* (Norton, 2002), and *hepatitis A* strains (Kim et al., 2016) in food and environmental samples, to name just a few examples. Another area in which real-time PCR is used is in industrial microbiology. An example is its use in the quantification of viable yeasts during wine fermentation and controlling the risk of wine spoilage (Hierro et al., 2006).

4.7.3 Clinical Oncology

4.7.3.1 Minimal Residual Disease

The detection of minimal residual disease (MRD) (i.e., the detection of a very low number of malignant cells) significantly correlates with the clinical outcome in many hematological malignancies. Therefore MRD monitoring is important for therapy guidance in clinical settings. Real-time PCR-based techniques for the detection of MRD have been developed and have proven to be more sensitive than classical morphological techniques, allowing for the detection of tumor cells at submicroscopic amounts (Kayser et al., 2015). Major challenges, however, which prevent its widespread use are the complexity and high cost of the assays (Rocha et al., 2016).

4.7.3.2 Single Nucleotide Polymorphisms

SNPs, present in the human genome, are used as genetic markers to follow the inheritance patterns of chromosomal regions from generation to generation. Moreover, they consist of a powerful tool in the study of genetic factors associated with human diseases (Johnson and Todd, 2000; Risch, 2000). Real -time PCR assays have been developed for SNP detection. For this application, the level of

discrimination between target and nontarget alleles is the most important challenge when optimizing the technique. In this real-time PCR method, one set of primers and two allele-specific fluorescent labeled probes are used, making use of different reporter dyes. The two alleles can thus be distinguished by the differential fluorescent emission of the two different reporter dyes (Livak, 1999). Alternatively, SNP detection often uses molecular beacons (Tyagi et al., 1998) or scorpion probes (Whitcombe et al., 1999). For each individual SNP assay, the level of discrimination has to be tested experimentally, since it is dependent on the mismatch.

4.7.3.3 Chromosomal Translocations

Chromosomal translocations, which often take place in tumor cells, can be employed as tumor-specific PCR targets. The primers are designed so that they anneal to opposite sides of the breakpoint in the fusion gene. These fusion genes are interesting targets for the design of a PCR method. Indeed, they are directly related to the oncogenic process and are therefore stable throughout the disease course. Moreover, the breakpoint fusion sites at the DNA level differ in each patient so that patient-specific real-time PCR strategies can be applied (Summers et al., 2001). Alternatively, real-time PCR can be performed on the transcripts of tumor-specific fusion genes. These are disease-specific transcripts located over chromosome breakpoints, leading to an in-frame RNA product (Rowley, 1998). Interestingly, these fusion transcripts can be identical in individual patients despite distinct breakpoints, because the breakpoints often are located in introns. An advantage of this approach is that the same set of primer/probes can be used for the analysis of individual patients with the same fusion transcript but different breakpoint translocations (Van Der Velden et al., 2003).

4.7.4 Gene Therapy

The primary goal of gene therapy is to specifically deliver the therapeutic gene to the target organ in a time- and dose-dependent manner and, most importantly, to avoid delivery to nontarget organs, since this may result in toxic side effects. In this case, the therapeutic gene is delivered to the target cells through a vector system, most often a viral vector (Crystal, 1995). Two important parameters have to be analyzed when considering gene therapy as a drug delivery system:

1. gene transfer estimation, which is the expression level of the therapeutic gene in regard to target tissue levels over time; and
2. biodistribution, which is the distribution of the drug in different organs for different routes of administration.

Real-time PCR assays have been validated in regard to gene transfer as well as biodistribution of gene therapy vectors. The major challenge in optimizing these assays is the accuracy of quantification and the sensitivity of the assay. Real-time PCR assays using hybridization probes, for instance, have been validated for adenovirus gene transfer vectors and proven to be quantitative, reproducible, and sensitive (Senoo et al., 2000; Hackett et al., 2000).

4.7.5 Validation of "-Omics" Related Technologies

The widespread use of several "-omics" techniques, such as microarrays, next generation sequencing, and epigenetic technologies, results in a massive amount of data. A confirmation of these results by an independent technique largely increases the validity of such data. qPCR plays an important role in such confirmation because of its detection sensitivity, sequence specificity, large dynamic range, and high precision and reproducible quantification. Disadvantages of qPCR (or RT-qPCR in this regard) are that it is labor intensive, time-consuming, and expensive. Moreover, several factors need to be taken into account that contribute to the correlation between microarray and qPCR results. Therefore often only a small part of the results are validated (Provenzano and Mocellin, 2007; Morey et al., 2006).

4.7.6 Forensic Genetics

DNA identifications have become standard use in forensic laboratories. Forensic laboratories are being confronted with an increasing number of DNA typing requests, paralleled by a decreasing amount of DNA per criminal case, which is urging them to develop qPCR assays that can simultaneously identify or quantify several parameters. Of importance for reliable qPCR results, next to the amount of total DNA or male DNA, is the quality of the DNA as well as the presence of inhibitory factors in the sample. As such, efforts by different labs have resulted in assay developments, in which one can simultaneously detect total human DNA, human male DNA, DNA degradation, and PCR inhibition (Hudlow et al., 2008; Ewing et al., 2016).

4.8 CRITERIA FOR OPTIMIZING QUANTITATIVE POLYMERASE CHAIN REACTION ASSAYS

The choice of method for performing qPCR reactions depends on the specific application envisioned. An important distinction has to be made, for instance, depending on whether one is dealing with measuring mRNA expression levels (RT-qPCR) or with DNA copy number (qPCR). Other factors, such as the number of genes, the number of

targets, the importance of fast screening, allelic discrimination, accurate quantification, sensitivity, and the costs, need careful consideration.

Probably one of the most widely used applications of qPCR is the quantification of mRNA expression. The use of RT-qPCR has been validated for numerous applications in fundamental research and clinical diagnostics. This is clear from the exponential increase in publications when performing a PubMed search using the words "reverse transcription" and "real-time PCR" (Fig. 4.3). Although the latter was developed at almost the same time as qPCR, it became rapidly clear that optimization of reliable RT-qPCR assays is much more complicated than qPCR assays. Indeed, reliable results depend on good RNA quality (Kirchner et al., 2014), consistent reverse transcriptase conditions (Bustin et al., 2015), the absence of coamplifying contaminating genomic DNA (Derveaux et al., 2010), and the appropriate normalization (Vandesompele et al., 2002). The latter refers to the correction for experimental variations in individual reverse transcriptase and PCR amplification efficiency. This is of major importance in RT-qPCR assays, since differences in efficiency of the reverse transcriptase reaction will result in an amount of cDNA that does not correspond to the starting amount of RNA. Furthermore, because of the exponential nature of the PCR reaction, minor differences in PCR amplification efficiency will result in major differences in the PCR product. Currently, the most widely applied method to correct for these variations is normalization to a reference gene. In theory, an ideal reference gene should be expressed at a constant level among different tissues of an organism, at all stages of development, and should not be affected by the experimental treatment itself. In practice, however, finding a gene with these characteristics is an almost impossible task. Indeed, the expression of a housekeeping gene can also be regulated by experimental treatment or can be tissue-dependent. It is therefore important that prior to performing each specific set of experiments, a suitable reference gene needs to be validated. For this purpose, one can test a range of widely used reference genes, such as β-actin, glyceraldehyde-3-phosphate-dehydrogenase, rRNA, hypoxanthine guanine phosphoribosyl transferase, ribosomal proteins, cyclophilin, mitochondrial ATP synthase 6, or porphobilinogen deaminase. Alternatively, different panels of reference genes are commercially available for testing. Of importance, one has to keep in mind that reference genes with similar functional properties may have a coregulated expression. It may therefore be advisable to choose reference genes from distinct functional classes. In 2002, Jo Vandesompele and coworkers published a method providing guidance and a solution on how to select for the most appropriate reference genes and to determine the minimum number of reference genes required in a specific experimental setting. They also provided a method for calculating a normalization factor based on the geometric mean of selected reference genes (Vandesompele et al., 2002).

Next to RT-qPCR for the quantification of mRNA expression levels, the field of RT-qPCR has received additional new interest with the discovery of noncoding RNAs, such as microRNA (Lee et al., 1993, 2003). MicroRNAs are small, noncoding RNAs of 20–22 base pairs that play a role in gene regulatory networks by binding to and repressing the activity of specific target mRNAs. In addition, microRNAs can be present in circulating body fluids, where they can serve as useful biomarkers for disease prediction or the follow-up of disease (He and Hannon, 2004). Because of their small size, the high homology between different miRNA family members, the absence of a common sequence feature such as a poly(A) tail, and the low abundance in body fluids, RT-qPCR quantification for miRNA requires specific adaptations associated with sample preparation, microRNA extraction and stabilization, experimental design, and data analysis (Mestdagh et al., 2014; Benes and Castoldi, 2010).

Although in theory the qPCR technique is probably the simplest technique in molecular biology, overall this contrasts largely with the way qPCR experiments in general, and RT-qPCR experiments in particular, are being performed. In a major attempt to standardize and improve the reliability of results reported in scientific publications based on this widespread technique, a group of international experts in the field of qPCR formulated and published the MIQE guidelines, referring to the "Minimum information required for the publication of qPCR Experiments" (Bustin et al., 2009). These guidelines describe the information that is considered to be essential in order to correctly perform and interpret qPCR experiments. With this, the goal of the authors was to ensure the integrity of scientific literature, promote consistency between different laboratories, and increase experimental transparency. Nine separate guidelines are formulated, related to the following items: (1) experimental design; (2) sample information; (3) nucleic acid extraction; (4) reverse transcription reaction; (5) gene target sequence; (6) oligonucleotides; (7) qPCR protocol; (8) the validation of qPCR; and (9) data analysis.

Altogether, if these guidelines are taken into account, one can be assured that the described protocol is detailed and clear so that all readers can correctly interpret or critically evaluate and implement (and if necessary repeat) the described experiments. Moreover, using these guidelines will encourage a better experimental practice, allowing for more reliable and unequivocal interpretation of qPCR results. With more than 3000 citations, the MIQE guidelines are widely recognized and accepted, thereby increasing the quality of reported qPCR results. But still, we are awaiting

an even more global implementation of the guidelines, as this would encourage detailed auditing of experimental details, data analysis, and reporting principles. General implementation is thus an important prerequisite for the further development of qPCR into a robust, accurate, and reliable nucleic acid quantification technology.

4.9 CONCLUSIONS

The introduction of qPCR technology has revolutionarily simplified the quantification of DNA and RNA. This has had a great impact in the field of molecular research and diagnostics, since enormous amounts of data can be obtained within a very short research time. The decreased costs for the thermal cyclers as well as for the necessary reagents for applying the technique have aided in its rapid increase in use. Indeed, its use increased exponentially since its development. Therefore it is clear that this technique has become the gold standard for the detection and quantification of DNA and RNA in general research or diagnostic laboratories.

Although qPCR assays by themselves are characterized by high precision and reproducibility, the accuracy of the data obtained are largely dependent on several other factors. It is insufficient to simply extend one's knowledge of classic end-point PCR to design and analyze experiments using qPCR. Indeed, many other controls are needed to be certain of the accuracy of the results when using qPCR assays. These factors, such as sample preparation, quality of the standard, choice of a relevant reference gene, and normalization of samples, need careful consideration and optimization. Furthermore, it is important that standardized criteria and international uniformity in experimental design and data analysis and reporting are reached to be able to compare data between different laboratories and correctly interpret the reliability of published data. The latter has improved greatly with the publication of the MIQE guidelines.

REFERENCES

Ahmad, A.I., 2007. BOXTO as a real-time thermal cycling reporter dye. J. Biosci. 32, 229–239.

Bailey, J.S., Cox, N.A., 1992. Universal preenrichment broth for the simultaneous detection of *Salmonella* and *Listeria* in foods. J. Food Prot.® 55, 256–259.

Benes, V., Castoldi, M., 2010. Expression profiling of microRNA using real-time quantitative PCR, how to use it and what is available. Methods 50, 244–249.

Bengtsson, M., Karlsson, H.J., Westman, G., Kubista, M., 2003. A new minor groove binding asymmetric cyanine reporter dye for real-time PCR. Nucleic Acids Res. 31, e45.

Bernard, P.S., Wittwer, C.T., 2000. Homogeneous amplification and variant detection by fluorescent hybridization probes. Clin. Chem. 46, 147–148.

Bhagwat, A.A., 2003. Simultaneous detection of Escherichia coli O157: H7, Listeria monocytogenes and Salmonella strains by real-time PCR. Int. J. Food Microbiol. 84, 217–224.

Bonetta, L., 2005. Prime time for real-time PCR. Nat. Methods 2, 305–312.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., Wittwer, C.T., 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin. Chem. 55, 611–622.

Bustin, S., Dhillon, H.S., Kirvell, S., Greenwood, C., Parker, M., Shipley, G.L., Nolan, T., 2015. Variability of the reverse transcription step: practical implications. Clin. Chem. 61, 202–212.

Cardullo, R.A., Agrawal, S., Flores, C., Zamecnik, P.C., Wolf, D.E., 1988. Detection of nucleic acid hybridization by nonradiative fluorescence resonance energy transfer. Proc. Natl. Acad. Sci. 85, 8790–8794.

Catrina, I.E., Marras, S.A., Bratu, D.P., 2012. Tiny molecular beacons: LNA/2′-O-methyl RNA chimeric probes for imaging dynamic mRNA processes in living cells. ACS Chem. Biol. 7, 1586–1595.

Clementi, M., Bagnarelli, P., Manzin, A., Menzo, S., 1994. Competitive polymerase chain reaction and analysis of viral activity at the molecular level. Genet. Anal. Biomol. Eng. 11, 1–6.

Crystal, R.G., 1995. Transfer of genes to humans: early lessons and obstacles to success. Science 270, 404–410.

De Crom, S.C., Rossen, J.W., Van Furth, A.M., Obihara, C.C., 2016. Enterovirus and parechovirus infection in children: a brief overview. Eur. J. Pediatr. 175 (8), 1023–1029.

Derveaux, S., Vandesompele, J., Hellemans, J., 2010. How to do successful gene expression analysis using real-time PCR. Methods 50, 227–230.

Endimiani, A., Jacobs, M.R., 2016. The changing role of the Clinical Microbiology Laboratory in defining resistance in Gram-negatives. Infect. Dis. Clin. North Am. 30, 323–345.

Ewing, M.M., Thompson, J.M., Mclaren, R.S., Purpero, V.M., Thomas, K.J., Dobrowski, P.A., Degroot, G.A., Romsos, E.L., Storts, D.R., 2016. Human DNA quantification and sample quality assessment: developmental validation of the PowerQuant system. Forensic Sci. Int. Genet. 23, 166–177.

Faloona, F., Mullis, K., Scharf, S., 1986. Specific enzymatic amplification of DNA in vitro. Cold Spring Harb. Symp. Quant. Biol. 263–273.

Gašparič, M.B., Cankar, K., Žel, J., Gruden, K., 2008. Comparison of different real-time PCR chemistries and their suitability for detection and quantification of genetically modified organisms. BMC Biotechnol. 8, 26.

Gibson, U., Heid, C.A., Williams, P.M., 1996. A novel method for real time quantitative RT-PCR. Genome Res. 6, 995–1001.

Giulietti, A., Overbergh, L., Valckx, D., Decallonne, B., Bouillon, R., Mathieu, C., 2001. An overview of real-time quantitative PCR: applications to quantify cytokine gene expression. Methods 25, 386–401.

Gray, J., Coupland, L., 2014. The increasing application of multiplex nucleic acid detection tests to the diagnosis of syndromic infections. Epidemiol. Infect. 142, 1–11.

Gudnason, H., Dufva, M., Bang, D.D., Wolff, A., 2007. Comparison of multiple DNA dyes for real-time PCR: effects of dye concentration and sequence composition on DNA amplification and melting temperature. Nucleic Acids Res. 35, e127.

Guyer, R.L., Koshland Jr., D.E., 1989. The molecule of the year. Science 246, 1543.

Hackett, N.R., El Sawy, T., Lee, L.Y., Silva, I., O'leary, J., Rosengart, T.K., Crystal, R.G., 2000. Use of quantitative TaqMan real-time PCR to track the time-dependent distribution of gene transfer vectors in vivo. Mol. Ther. 2, 649−656.

Hartnell, R., Lowther, J., Avant, J., Dancer, D., Lees, D., Russell, J., 2012. The development of LENTICULES as reference materials for noroviruses. J. Appl. Microbiol. 112, 338−345.

He, L., Hannon, G.J., 2004. MicroRNAs: small RNAs with a big role in gene regulation. Nat. Rev. Genet. 5, 522−531.

Heid, C.A., Stevens, J., Livak, K.J., Williams, P.M., 1996. Real time quantitative PCR. Genome Res. 6, 986−994.

Heller, M.J., Morrison, L.E., 1985. Chemiluminescent and fluorescent probes for DNA hybridization systems. Rapid Detect. Identif. Infect. agents 245−256.

Hierro, N., Esteve-Zarzoso, B., Gonzalez, A., Mas, A., Guillamon, J.M., 2006. Real-time quantitative PCR (QPCR) and reverse transcription-QPCR for detection and enumeration of total yeasts in wine. Appl. Environ. Microbiol. 72, 7148−7155.

Higuchi, R., Fockler, C., Dollinger, G., Watson, R., 1993. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Biotechnology 11, 1026−1030.

Holland, P.M., Abramson, R.D., Watson, R., Gelfand, D.H., 1991. Detection of specific polymerase chain reaction product by utilizing the 5'—3'exonuclease activity of Thermus aquaticus DNA polymerase. Proc. Natl. Acad. Sci. 88, 7276−7280.

Hudlow, W.R., Chong, M.D., Swango, K.L., Timken, M.D., Buoncristiani, M.R., 2008. A quadruplex real-time qPCR assay for the simultaneous assessment of total human DNA, human male DNA, DNA degradation and the presence of PCR inhibitors in forensic samples: a diagnostic tool for STR typing. Forensic Sci. Int. Genet. 2, 108−125.

Johnson, G.C., Todd, J.A., 2000. Strategies in complex disease mapping. Curr. Opin. Genet. Dev. 10, 330−334.

Kayser, S., Walter, R.B., Stock, W., Schlenk, R.F., 2015. Minimal residual disease in acute myeloid leukemia−current status and future perspectives. Curr. Hematol. Malig. Rep. 10, 132−144.

Kim, M.J., Lee, S.Y., Kim, H.J., Lee, J.S., Joo, I.S., Kwak, H.S., Kim, H.Y., 2016. Development of a one-step duplex RT-PCR method for the simultaneous detection of VP3/VP1 and VP1/P2B regions of the hepatitis A virus. J. Microbiol. Biotechnol.

Kirchner, B., Paul, V., Riedmaier, I., Pfaffl, M.W., 2014. mRNA and microRNA purity and integrity: the key to success in expression profiling. Methods Mol. Biol. 1160, 43−53.

Kutyavin, I.V., Afonina, I.A., Mills, A., Gorn, V.V., Lukhtanov, E.A., Belousov, E.S., Singer, M.J., Walburger, D.K., Lokhov, S.G., Gall, A.A., 2000. 3′-minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. Nucleic Acids Res. 28, 655−661.

Lee, M., Siddle, A., Page, R., 2002. ResonSense®: simple linear fluorescent probes for quantitative homogeneous rapid polymerase chain reaction. Anal. Chim. Acta 457, 61−70.

Lee, R.C., Feinbaum, R.L., Ambros, V., 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. cell 75, 843−854.

Lewis, B.P., Shih, I.-H., Jones-Rhoades, M.W., Bartel, D.P., Burge, C.B., 2003. Prediction of mammalian microRNA targets. Cell 115, 787−798.

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2—ΔΔCT method. Methods 25, 402−408.

Livak, K.J., 1999. Allelic discrimination using fluorogenic probes and the 5′ nuclease assay. Genet. Anal. Biomol. Eng. 14, 143−149.

Mestdagh, P., Hartmann, N., Baeriswyl, L., Andreasen, D., Bernard, N., Chen, C., Cheo, D., D'andrade, P., Demayo, M., Dennis, L., 2014. Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study. Nat. Methods 11, 809−815.

Moreau, V., Voirin, E., Paris, C., Kotera, M., Nothisen, M., Remy, J.-S., Behr, J.-P., Erbacher, P., Lenne-Samuel, N., 2009. Zip Nucleic Acids: new high affinity oligonucleotides as potent primers for PCR and reverse transcription. Nucleic Acids Res. 37 (19). http://dx.doi.org/10.1093/nar/gkp661.

Morey, J.S., Ryan, J.C., Van Dolah, F.M., 2006. Microarray validation: factors influencing correlation between oligonucleotide microarrays and real-time PCR. Biol. Proced. Online 8, 175−193.

Nazarenko, I.A., Bhatnagar, S., Hohman, R., 1997. A closed tube format for amplification and detection of DNA based on energy transfer. Nucleic Acids Res. 25, 2516−2521.

Nazarenko, I., Lowe, B., Darfler, M., Ikonomi, P., Schuster, D., Rashtchian, A., 2002. Multiplex quantitative PCR using self-quenched primers labeled with a single fluorophore. Nucleic Acids Res. 30, e37.

Nielsen, P.E., Egholm, M., Berg, R.H., Buchardt, O., 1991. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. Science 254, 1497−1500.

Norton, D.M., 2002. Polymerase chain reaction-based methods for detection of Listeria monocytogenes: toward real-time screening for food and environmental samples. J. AOAC Int. 85, 505−515.

Paris, C., Moreau, V., Deglane, G., Voirin, E., Erbacher, P., Lenne-Samuel, N., 2010. Zip nucleic acids are potent hydrolysis probes for quantitative PCR. Nucleic Acids Res. 38, e95.

Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT−PCR. Nucleic Acids Res. 29, e45.

Provenzano, M., Mocellin, S., 2007. Complementary techniques: validation of gene expression data by quantitative real time PCR. Adv. Exp. Med. Biol. 593, 66−73.

Rijpens, N.P., Herman, L.M., 2002. Molecular methods for identification and detection of bacterial food pathogens. J. AOAC Int. 85, 984−995.

Ririe, K.M., Rasmussen, R.P., Wittwer, C.T., 1997. Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. Anal. Biochem. 245, 154−160.

Risch, N.J., 2000. Searching for genetic determinants in the new millennium. Nature 405, 847−856.

Rocha, J.M., Xavier, S.G., De Lima Souza, M.E., Assumpcao, J.G., Murao, M., De Oliveira, B.M., 2016. Current strategies for the detection of minimal residual disease in Childhood acute Lymphoblastic leukemia. Mediterr. J. Hematol. Infect. Dis. 8, e2016024.

Rowley, J.D., 1998. The critical role of chromosome translocations in human leukemias. Annu. Rev. Genet. 32, 495−519.

Rudi, K., Nogva, H.K., Moen, B., Nissen, H., Bredholt, S., Moretro, T., Naterstad, K., Holck, A., 2002. Development and application of new nucleic acid-based technologies for microbial community analyses in foods. Int. J. Food Microbiol. 78, 171−180.

Saitta, E., Brasselet, A., Cadouot, M., Gibeaud, A., Jourdain, M., Bozonnet, D., Goussot, V., Riedinger, J.M., Lizard, S., 2016. Validation steps for the detection of L858R mutation in EGFR gene by real time quantitative PCR. Ann. Biol. Clin. Paris. 74, 196−202.

Senoo, M., Matsubara, Y., Fujii, K., Nagasaki, Y., Hiratsuka, M., Kure, S., Uehara, S., Okamura, K., Yajima, A., Narisawa, K., 2000. Adenovirus-mediated in utero gene transfer in mice and guinea pigs: tissue

distribution of recombinant adenovirus determined by quantitative TaqMan–polymerase chain reaction assay. Mol. Genet. Metab. 69, 269–276.

Stappers, M.H., Hagen, F., Reimnitz, P., Mouton, J.W., Meis, J.F., Gyssens, I.C., 2016. Bacteroides fragilis in biopsies of patients with major abscesses and diabetic foot infections: direct molecular versus culture-based detection. Diagn Microbiol. Infect. Dis. 85, 263–265.

Stryer, L., 1978. Fluorescence energy transfer as a spectroscopic ruler. Annu. Rev. Biochem. 47, 819–846.

Summers, K.E., Goff, L.K., Wilson, A.G., Gupta, R.K., Lister, T.A., Fitzgibbon, J., 2001. Frequency of the Bcl-2/IgH rearrangement in normal individuals: implications for the monitoring of disease in patients with follicular lymphoma. J. Clin. Oncol. 19, 420–424.

Svanvik, N., Westman, G., Wang, D., Kubista, M., 2000. Light-up probes: thiazole orange-conjugated peptide nucleic acid for detection of target nucleic acid in homogeneous solution. Anal. Biochem. 281, 26–35.

Thelwell, N., Millington, S., Solinas, A., Booth, J., Brown, T., 2000. Mode of action and application of Scorpion primers to mutation detection. Nucleic Acids Res. 28, 3752–3761.

Tyagi, S., Kramer, F.R., 1996. Molecular beacons: probes that fluoresce upon hybridization. Nat. Biotechnol. 14, 303–308.

Tyagi, S., Bratu, D.P., Kramer, F.R., 1998. Multicolor molecular beacons for allele discrimination. Nat. Biotechnol. 16, 49–53.

Van Der Velden, V., Hochhaus, A., Cazzaniga, G., Szczepanski, T., Gabert, J., Van Dongen, J., 2003. Detection of minimal residual disease in hematologic malignancies by real-time quantitative PCR: principles, approaches, and laboratory aspects. Leukemia 17, 1013–1034.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., Speleman, F., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3 (7) research0034.

Von Ahsen, N., Oellerich, M., Schütz, E., 2000. Use of two reporter dyes without interference in a single-tube rapid-cycle PCR: α1-antitrypsin genotyping by multiplex real-time fluorescence PCR with the LightCycler. Clin. Chem. 46, 156–161.

Warhurst, G., Dunn, G., Chadwick, P., Blackwood, B., Mcauley, D., Perkins, G.D., Mcmullan, R., Gates, S., Bentley, A., Young, D., Carlson, G.L., Dark, P., 2015. Rapid detection of health-care-associated bloodstream infection in critical care using multipathogen real-time polymerase chain reaction technology: a diagnostic accuracy study and systematic review. Health Technol. Assess. 19, 1–142.

Whitcombe, D., Theaker, J., Guy, S.P., Brown, T., Little, S., 1999. Detection of PCR products using self-probing amplicons and fluorescence. Nat. Biotechnol. 17, 804–807.

Wittwer, C.T., Herrmann, M.G., Gundry, C.N., Elenitoba-Johnson, K.S., 2001. Real-time multiplex PCR assays. Methods 25, 430–442.

Zipper, H., Brunner, H., Bernhagen, J., Vitzthum, F., 2004. Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications. Nucleic Acids Res. 32, e103.

Chapter 5

Integrated Polymerase Chain Reaction Technologies (Sample-to-Answer Technologies)

P.B. Danielson[1,2], H.E. McKiernan[2] and K.M. Legg[2]
[1]*University of Denver, Denver, CO, United States;* [2]*Center for Forensic Science Research and Education, Willow Grove, PA, United States*

5.1 INTRODUCTION

5.1.1 Standard Workflow for Polymerase Chain Reaction-Based Assays

The tools available for infectious disease diagnostics and human identity testing have benefitted from major technological advances. Among the most significant of these advances was the invention of the polymerase chain reaction (PCR) in the mid-1980s (Saiki et al., 1988). This made it possible to easily and quickly amplify specific nucleic acid sequences from microscopic quantities of DNA collected from a patient specimen or recovered from an item of evidence in a criminal case.

Regardless of the type of specimen being analyzed, however, the first step in any PCR-based molecular genetic assay was the preparation of a nucleic acid extract. While this was originally performed manually (typically by boiling or chemical lysis, that is, the use of noxious organic solvents like phenol—chloroform and ethanol precipitation), a wide variety of commercially available instruments have since been developed that automate this process (Fujii et al., 2013). Automated cell lysis can be achieved through the use of detergents and chaotropic salts, heat, mechanical disruption, or simple pressure. Automated nucleic acid extraction/purification typically employs some variant of solid phase extraction whereby nucleic acids are selectively but reversibly bound to a solid substrate (typically silica), thereby allowing protein and other unwanted cellular components to be washed away, after which the now pure nucleic acids are eluted into an appropriate buffer for downstream applications. Such automated platforms provide faster and more consistent results than manual methods. The use of 96-well, 384-well, and higher density specimen processing formats has also greatly increased specimen throughput, which has led to improved economies of scale.

The second step is the PCR amplification process, which, since the 1980s, has been performed by programmable thermal cyclers. By means of heating blocks or temperature-controlled air chambers, specimens cycle through repeated rounds of DNA denaturation, target-specific primer annealing, and nascent strand extension. Here, technical advances have focused on polymerase functionalities (e.g., processivity, fidelity, stability), reduced reaction volumes, and improved thermal transitions to reduce the total amplification run times.

The third step is the detection and analysis of the amplified products of the PCR. Size or conformational fractionation by slab-gel or capillary electrophoresis (CE) has long been the default technology. Here, the availability of precast gels, capillary arrays connected to gel pumps, and higher-density formats for automated sample loading has helped to increase specimen throughput. However, the process remains somewhat laborious and time-consuming. Advances in alternative technologies for PCR amplicon detection and analysis have sought to circumvent the limitations of electrophoresis-based approaches by detecting target amplicons during the amplification process itself. These approaches employ intercalating compounds, minor-groove binding dyes, and a wide variety of hybridization probes (e.g., molecular beacons and Scorpions®) that can be detected and quantified as they bind to nascent strands of DNA created during each cycle of the PCR process.

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00005-5
Copyright © 2017 Elsevier Ltd. All rights reserved.

Another approach that avoids the need for post-amplification fractionation by gel electrophoresis is melt curve analysis. This is a postamplification assay that can be performed without the need to remove an aliquot of amplified DNA from the PCR tube. It examines the helix-dissociation profile (i.e., the melt curve) of nascent double-stranded DNA as it is subjected to increasing temperatures. The melt curve of any amplicon is a function of sequence-specific nearest-neighbor thermodynamic interactions. As a result, melt curves serve as "signatures" for the presence of an amplified target sequence. See Chapters 4 and 6 for a comprehensive discussion of the underlying principles of real-time PCR, high-resolution melt curve analysis, and their respective applications to molecular diagnostics.

The final step in PCR-based molecular assays is data interpretation, whereby the fragments detected by gel electrophoresis or CE, the real-time PCR amplification curves, and the postamplification melt curves are interpreted. For narrowly tailored and well-optimized PCR assays with clearly defined expected outcomes, interpretation can be a relatively simple and rapid process. For example, PCR-based diagnostic tests for methicillin-resistant *Staphylococcus aureus*, *Clostridium difficile*, and tuberculosis (TB) are relatively simple, low-multiplexed assays that yield well-defined clinical answers. In sharp contrast to this are PCR-based assays for human identity testing, which may be used for kinship analysis or in forensic DNA crime laboratories. The use of short tandem repeats (STRs) as target genetic markers and the highly variable nature of forensic evidence with regard to sample integrity, DNA quantity, and contributor complexity make data interpretation an especially time- and labor-intensive process, requiring manual review by a skilled laboratory analyst. Efforts to streamline this process have focused on the development of powerful expert systems (computer software that is guided by user-defined rules with the goal of being able to consistently analyze a given set of data with the same skill as a human) (Haned and Gill, 2015). Expert systems do not replace humans. Rather, they automate tedious and often time-consuming aspects of data analysis to provide analysts with quality scoring of analyzed data and explanations for the reasoning that the program uses to support a specific analytical outcome. See Chapter 21 for a comprehensive discussion of the use of DNA testing for human identity applications.

Even with the advances that have been made in laboratory automation, the traditional diagnostic assay workflow has remained largely segmented. Specimens and reagents are usually manually loaded and unloaded from different instruments and transported among different physical locations within a laboratory. At a minimum, these include dedicated pre-PCR (low-template extraction and purification) and a post-PCR (high-template amplification and postamplification analysis) areas of the laboratory. The purpose for this segregation of work areas is to minimize the chance of generating erroneous assay results (the inadvertent introduction of previously amplified nucleic acids into preamplification nucleic acid extracts or preamplification processes). In both a medical diagnostic context and a forensic context, undetected contamination of one specimen with PCR-amplified products from a different specimen could have catastrophic consequences.

It is widely recognized that even with stringent standard operating procedures and the utmost caution, it is impossible to completely eliminate the risk of cross-contamination when performing PCR-based assays using segmented workflows, even if the individual segments are automated. For this reason, laboratories typically employ multiple layers of controls and quality checks that are designed not only to prevent cross-contamination but also to ensure that it is detected in those rare instances when it does occur. Subsequent root cause analysis and corrective action reports are then generated in an effort to perpetually improve the overall process.

5.1.2 Fully Integrated Polymerase Chain Reaction-Based Assay Systems

There has been a remarkable shift from labor-intensive assays and segmented workflows to fully automated and integrated instruments. Where traditional DNA-based assays require extensive hands-on time to isolate nucleic acids from test specimens; amplify diagnostically informative amplicons; and convert raw electrophoretic fractionation, high-resolution melt curve, or hybridization data into diagnostically meaningful results, a new generation of easy to use, fully automated assays has greatly streamlined this entire process. Variously described as "sample-to-answer" or "sample-in, answer-out" systems in medical diagnostics or as "sample-to profile" systems in human identity testing, these instruments integrate and fully automate nucleic acid extraction, amplification, detection, data interpretation, and reporting in a single device (Park et al., 2011; Sackmann et al., 2014; Buchan and Ledeboer, 2014).

The miniaturization of instruments and reductions in reaction volumes without a loss of data output quality have been hallmarks of modern molecular biology, and these advances are at the core of the sample-to-answer assays. All necessary chemical reagents for the test specimens and some form of internal process control are prepackaged in either a liquid or dried state in disposable assay cassettes or blister packs. Typically, each specimen is processed along its own dedicated pathway, which physically isolates it from other specimens during the entire analytical process. This eliminates the opportunities for potential cross-contamination that exist in traditional segmented workflows. As a result, laboratories are able to bypass many of the burdensome and expensive procedural and infrastructural aspects of

contamination control. Most of these systems are lightweight benchtop instruments with small footprints that make them portable and thus well suited to near point-of-care diagnostics, even in lower resource environments. Similarly, sample-to-profile human identity testing systems are compact enough to be used in a police department booking station, mobile crime scene van, airport security area, or field-forward military site.

These instruments and their assay cassettes can be run with minimal hands-on operation and by staff with less technical training than is typically needed for a traditional diagnostic lab technologist or forensic analyst. By eliminating segmented workflows and assaying specimens in parallel, these systems typically have much faster turnaround times than conventional assays. In the medical diagnostic arena, the sensitivity and specificity of sample-to-answer assay systems are comparable to that of more conventional methods for the identification of the causative organism responsible for a given pathology and often times the presence in that organism of specific genes associated with antimicrobial resistance traits and/or virulence factors. Similarly, the accuracy of sample-to-profile instruments for human identity testing is comparable to that of conventional methods for the analysis of buccal swabs and other good quality reference type samples.

5.2 COMMERCIAL SAMPLE-TO-ANSWER ASSAY SYSTEMS

5.2.1 Systems for Infectious Disease Diagnostics

A number of sample-to-answer and sample-to-profile systems are commercially available (Table 5.1), including the FilmArray® (BioFire® Diagnostics/bioMérieux), the GeneXpert® (Cepheid), the Liat® PCR system (cobas®/Roche), Simplexa™ for 3M Integrated Cycler (Focus Diagnostics), the Verigene® System (Nanosphere/Luminex®), the RapidHIT™ 200 DNA profiler (IntegenX), and the DNAscan™ Rapid DNA Analysis System (NetBio/General Electric [GE] Healthcare). While certainly not an exhaustive list, these examples encompass the range of general operational principles underlying these systems. Specific details in regard to the individual assay chemistries that run on these systems will be provided in this chapter in the context of specific clinical or identity testing applications.

TABLE 5.1 Commercially Available Sample-to-Answer Assay Systems

System	Manufacturer (website)	Application	Operating principle/Detection principle	Samples/run	Approximate run time
FilmArray®	BioFire® Diagnostics/bioMérieux (www.biofiredx.com)	Infectious disease diagnostics	Nested NAAT Melt curve analysis	1	70 min
GeneXpert®	Cepheid (www.cepheid.com/us)	Infectious disease diagnostics	RT-PCR NAAT Fluorescent molecular beacon	1–80	30–150 min
Liat®	cobas®/Roche (www.usdiagnostics.roche.com/en/instrument/cobas-liat.html)	Infectious disease diagnostics	RT-PCR NAAT TaqMan® probe	1	20 min
Simplexa™ 3M Integrated Cycler	Focus Diagnostics (www.focusdx.com/product-catalog/simplexa)	Infectious disease diagnostics	RT-PCR NAAT Fluorescent probes	8	60 min
Verigene®	Nanosphere/Luminex® (www.nanosphere.us)	Infectious disease diagnostics	Microarray capture direct hybridization Light scattering by derivitized gold nanoparticles	1	150 min
DNAscan™	NetBio/GE Healthcare (www.gelifesciences.com)	Human identity testing	Multiplex PCR Microfluidic CE Electrophoresis	5	90 min
RapidHit™ 200	IntegenX (www.integenx.com)	Human identity testing	Multiplex PCR integrated conventional CE	1–7	90 min

The FilmArray® system (Fig. 5.1) provides a generic illustration of a fully integrated PCR-based sample-to-answer system. The FilmArray® system combines an automated in vitro diagnostic instrument (the base unit) with assay-specific "pouches" to detect multiple nucleic acid targets in clinical specimens (Xu et al., 2013). Each assay pouch is a disposable, self-contained, closed system that contains all the reagents required to extract, amplify, and detect specific nucleic acid targets that may be present in a clinical specimen (Poritz et al., 2011). The reservoirs in the rigid plastic component (the fitment) of the pouch contain lyophilized reagents. The flexible plastic film portion of the pouch is divided into a series of blister pack-like compartments. A series of chemical processes are executed through interactions between the pouch and actuators and sensors in the base unit. These include cell lysis and nucleic acid extraction and purification from the test specimen; first-stage multiplex PCR; and second-stage singleplex PCR and melting analysis, both of which take place in a multiwell array. To run the assay, a technician uses a syringe to load "Hydration Solution" into the pouch. This rehydrates the lyophilized reagents that are prepackaged in the pouch fitment. A patient specimen is then mixed with "Sample Buffer," which inactivates RNases in the specimen and which will later facilitate binding of the nucleic acids to the magnetic beads for nucleic acid isolation. A sample loading syringe is then used to transfer the specimen/sample buffer mixture to the pouch. The pouch is then loaded into the base unit.

After the technician initiates a specimen run, a series of plungers, pneumatic actuators, and hard seals work together to mix the liquid reagents and move them between the blisters of the pouch (Grover et al., 2003). The FilmArray® base unit automatically performs these functions based on the run protocol selected for a specific pouch (assay chemistry) and specimen type in the instrument's operating software. Nucleic acid isolation occurs in the first three blister compartments of the pouch. The first step in processing a specimen is to lyse the outer membrane of the target microbes that may be in the patient specimen. This is done using a device called a bead-beater. A sensor in the base unit monitors the speed and operation of the bead-beater and aborts the run if the bead-beater is not working properly. Following bead-beating, the nucleic acids contained in the sample are captured, washed, and eluted by magnetic bead technology. A retractable magnet is used to capture or release the magnetic beads during washes. These steps require about 10 min to complete.

The next step in the process is reverse transcription (RT) and first-stage multiplex PCR amplification. The RT

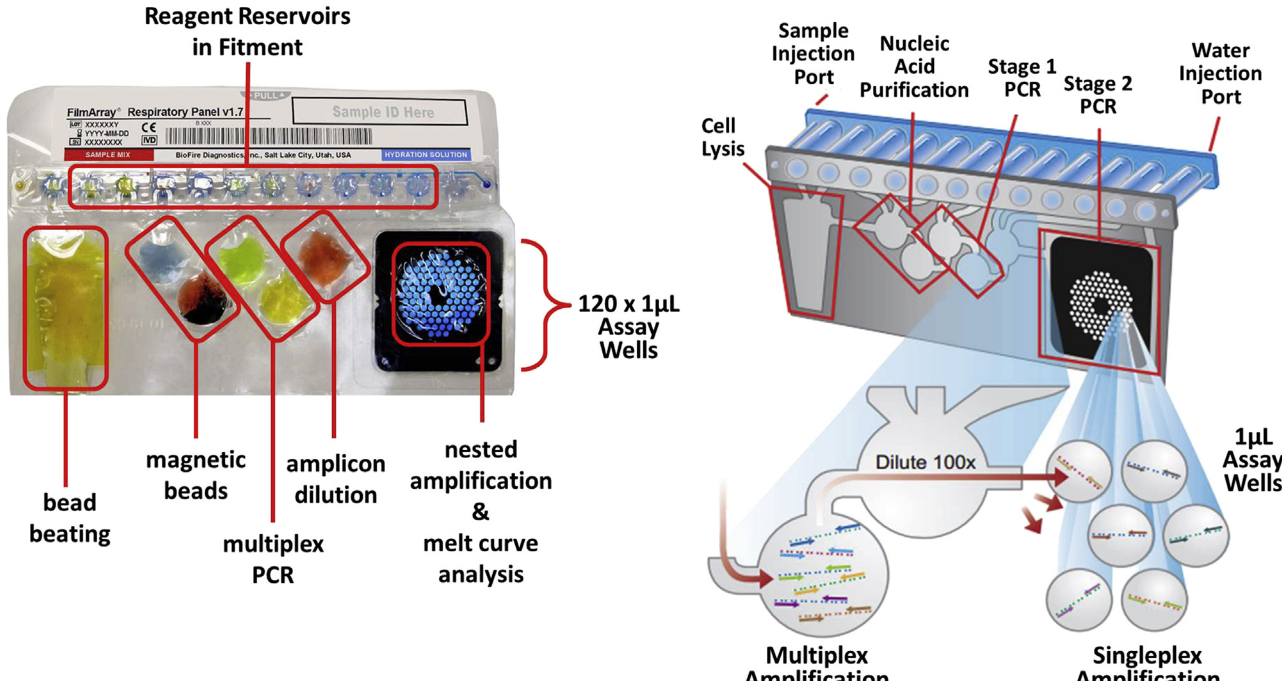


FIGURE 5.1 Photographic image (left) and schematic diagram (right) showing the salient features of a FilmArray® assay system consisting of a hard plastic fitment containing reagent reservoirs and a flexible plastic film pouch consisting of a series of reaction chambers and transfer channels. Once a sample has been introduced through the sample injection port, the pouch is inserted into a FilmArray® instrument base unit (not shown) for processing. The colored liquids shown in the photograph are to facilitate visualization. Actual FilmArray® pouches do not contain colored liquids. *PCR*, polymerase chain reaction. *Images adapted from FilmArray® 2.0 Operators Manual CE IVD.*

step is necessary to convert RNA into cDNA prior to first-stage multiplex PCR amplification in those assays where target pathogens for identification have RNA genomes (e.g., coronavirus, orthomyxovirus, picornavirus, norovirus, and sapovirus). The purified nucleic acid extract is mixed with a preheated master mix to initiate the RT step, and subsequent thermocycling for multiplex PCR amplifies all of the targets identified by the pouch as well as the process control amplicon. A Peltier device drives the thermocycling both for the RT reaction (when required) and for the first-stage PCR reactions. The purpose of the first-stage PCR is to enrich for the target nucleic acids if they are present in the patient specimen.

The amplified products of first-stage PCR are then diluted and combined with fresh PCR reagents containing a saturating fluorescent intercalating DNA dye (LCGreen® Plus). This solution is distributed to a second-stage PCR array. Each of the individual wells of the array contain primers for different assays (in triplicate) that target specific nucleic acid sequences from each of the pathogens targeted by the pouch assay as well as primers for the internal positive control(s). The annealing sites for the second-stage PCR primers are located within the sequence of the amplified products generated during the first-stage multiplex PCR (i.e., nested PCR). This serves to enhance both the sensitivity and specificity of the assay.

After completion of the last cycle of the second-stage PCR, a DNA melting curve analysis is performed to identify positive and negative PCR reactions. This is used to determine which of the targeted microbes were present in the patient specimen. During the second-stage PCR process, LCGreen® Plus dye is incorporated into the copies of DNA as they are made during each PCR cycle. When bound to double-stranded DNA, the dye fluoresces. As the temperature is increased and the copies of double-stranded DNA melt, the dye is released with a corresponding drop in fluorescence. A DNA melting curve is generated by slowly increasing the temperature of the PCR array. As this is done, the fluorescence emitted by the DNA-bound dye in each well of the array is imaged by a camera and recorded. A second Peltier device controls thermocycling for second-stage PCR and for the DNA melt analysis. As with the first-stage PCR, the thermocycling profile is controlled in accordance with the programmed run protocol for the specific reagent pouch assay being performed. Both the second-stage PCR and the melt analysis take place in the array located in the final pouch blister. Because the melting profile for the target amplicon for each microbe in the assay is known, PCR products from specific microbes can be readily identified.

Two internal positive controls are incorporated into the assay. First there is an "RNA Process Control" (RPC). The RPC employs lyophilized yeast (*Schizosaccharomyces pombe*), which is prepackaged in the assay pouch and rehydrated when the patient specimen is loaded. The yeast is carried through all stages of the test process, including lysis, nucleic acid purification, the first-stage PCR, dilution, the second-stage PCR, and the DNA melting curve analysis. An RNA transcript present in the RPC is targeted for amplification, so a positive control result indicates that all steps carried out in the assay pouch were successful. A second-stage PCR (PCR2) control is also included. The PCR2 control detects a DNA target that is dried into wells of the array along with corresponding amplification primers. A positive result indicates the success of the second-stage PCR.

The FilmArray® base unit's software automatically analyzes the melt curve data from the replicate wells for each second-stage PCR reaction (both controls and test samples) and automatically generates a test report at the end of the run indicating which of the targeted microbes were detected. If either of the controls fail, all results will be reported as "Invalid." Based on independent workflow analysis studies (Butt et al., 2014), the entire process from specimen-to-answer takes just over 1 h to complete but only about 5 min of actual hands-on time.

Other commercially available platforms take a similar approach by integrating specimen processing in a single-use cartridge system, though the specifics of nucleic acid extraction, target amplification, and detection vary. The GeneXpert® system, for example, combines a base unit analyzer with small disposable cartridges (modules) that are barcoded and preloaded with all of the necessary reagents for particular assays (Lawn et al., 2013; Tortoli et al., 2012). A patient specimen is placed into the cartridge, which is then loaded into the base unit, which reads the cartridge barcode and initiates an assay-specific protocol. Nucleic acid extraction takes place in a processing chamber that contains reagents, filters, and capture technologies necessary to extract, purify, and amplify target nucleic acids. Nucleic acid extraction employs a combination of chemical and ultrasonic lysis, followed by RT-PCR and fluorescent monitoring of multiple channels for target and control signals. Integrated valves are used to facilitate fluid transfer from chamber to chamber, and thin PCR chamber walls enable rapid thermal cycling. The systems software interprets the detected signal and generates a report of the results with a total run time of about 1 h. This system's base units are available with 1-, 2-, 4-, 16-, or 80-module configurations and are equipped with a robotic loading system to minimize hands-on time even in higher volume laboratories. Each module within a base unit operates individually, and as many different assays (e.g., *Mycobacterium tuberculosis*, methicillin-resistant *S. aureus*, or *C. difficile*) as necessary can be run by a technician at the same time. Therefore this system is both flexible and scalable to the needs of the specific laboratory. This is a key advantage relative to the FilmArray® system and other systems that process one assay pouch at a time. FilmArray®, however, is

able to screen for a greater number (up to 20) of target pathogens per pouch.

The Liat™ which stands for "Lab In A Tube" system is another real-time PCR-based assay system that utilizes small disposable linear tubes (Binnicker et al., 2015; Chen et al., 2015). Each tube contains all of the reagents necessary to perform an assay. These reagents are arranged in a series of blister pack compartments. Using a transfer pipette, a patient specimen is introduced to the top of the assay tube. After the assay tube is capped, it is inserted into an analyzer base unit, which integrates all nucleic acid testing processes, including reagent preparation, target enrichment, inhibitor removal, nucleic acid extraction, amplification, and real-time detection. Specimen sampling and handling are controlled using multiple sample processing modules contained within the base unit. The specimen processing modules consist of two assemblies. The first is a moving side assembly comprised of multiple sample processing plungers and clamps. The second is a fixed-side assembly. The plungers and clamps selectively compress the Liat™ assay tube segments against the fixed-side assembly so as to selectively release reagents from individual tube blister compartments and move the sample from one compartment to another. The nucleic acid extraction process is based on lysis by chaotropic salts followed by solid phase magnetic particle-based nucleic acid purification. The detection of amplified targets employs fluorogenically labeled hydrolysis probes (e.g., *Taq*Man® probes). As with other integrated systems, an internal positive control is used to confirm the instrument performance and result determination. The internal positive control (IPC) comprises an encapsulated RNA that is prepacked in each Liat™ tube. The base analyzer unit controls reaction conditions, such as thermal cycling temperatures, in accordance with specific assay programs. Data analysis employs predefined decision algorithms to generate a report of assay results. Like the FilmArray® system, the Liat™ system processes only one specimen at a time, but results are typically generated in only 20 min.

The Simplexa™ Molecular Assay system uses disk-based centrifugal microfluidics and real-time PCR assays, which are designed to run on the 3M™ Integrated Cycler (Strohmeier et al., 2015; Miao et al., 2015). Users are given a choice of two operational modes. In the first mode, a thermocycle-only disk is used, which allows for up to 96 standard real-time PCR assays but requires that patient specimens be extracted separately so that the disk can be loaded with purified DNA. Each of 96 radially inward-oriented inlet wells is connected to one of 96 amplification wells positioned at the outer rim of the disk. Contact heating is employed for thermocycling, and real-time detection can record data in up to four fluorescence channels. In the second mode of operation, a direct amplification disk is used to provide true sample-to-answer functionality. In this second mode, space on the direct amplification disk is occupied mostly by extraction microfluidics, and therefore only eight fully integrated sample processing reactions can be run at a time. The use of the disk requires little hands-on time aside from pipetting 50 μL of reaction mix onto the disk prior to processing a 50-μL patient sample. Samples to be assayed are pipetted into wells near the center of the disk. The technician then enters the assay run parameters and initiates the run. As the disk spins, centrifugal force moves the patient specimens and reaction reagents through the reaction chambers. The direct amplification disk makes use of direct amplification chemistries that can perform nucleic acid extraction and amplification in one protocol. Infrared energy and a high-velocity fan are used to heat and cool the real-time PCR reaction chambers on the disk. Rapid temperature transitions and efficient heat transfers translate into shorter cycle times (as little as 20 s/cycle). As with the thermocycle-only disk, up to four fluorescence channels are available for target detection, and the system software collects and analyzes the results. The Simplexa™ system provides results from a patient swab in about 1 h.

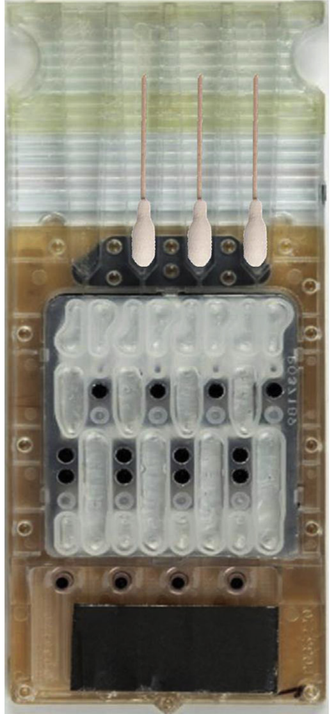
A similar specimen processing approach with a quite different detection technology is employed by the Verigene® system, which uses self-contained test cartridges in conjunction with two separate instrument modules: a cartridge processor and a cartridge reader. Each disposable test cartridge is designed for the multiplex analysis of a single patient specimen. It features a microfluidic cassette that contains all of the hybridization reagents needed for the assay and captures all of the waste materials that are generated in the process. It also contains a glass slide that serves as a solid support for the microarray where any targeted nucleic acids can be captured for detection. The cartridge processor unit manages the automated nucleic acid extraction, purification, amplification (if required), and hybridization processes. In the test cartridge, genomic DNA is extracted and sheared by sonication into 300 to 500 base pair fragments. This fragment size is easily manipulated and can be readily hybridized to other molecules. The fragmented genomic DNA is allowed to hybridize to a microarray of capture probes that are attached to the glass slide in the assay cartridge. Oligonucleotide probes conjugated to gold nanoparticles are then introduced and allowed to bind to any complementary genomic DNA fragment present on the microarray. Any unbound oligonucleotide-gold nanoparticle probes are washed away while elemental silver is deposited onto the gold nanoparticle probes, which are bound and remain. This serves to amply the signal for the optical detection of light scattering from the derivitized gold nanoparticles (Giljohann et al., 2010). The detection sensitivity that is achieved through the

use of gold nanoparticles exceeds that of the fluorescent dye-based detection methods that are employed by most other sample-to-answer systems, so much so that while nucleic acid amplification is an option with the Verigene® system, it is not always necessary and can be omitted in some assays. Finally, the cartridge reader images the microarray and analyzes the results from the processed cartridges in order to generate a report. Based on independent evaluations of this system (Butt et al., 2014), results are typically generated in approximately 2.5 h, making the system competitive with other sample-to-answer systems. The total amount of hands-on time was 21 min. The longer hands-on time was attributed to the need to thaw frozen reagents and move cartridges between the analyzer and the reader.

5.2.2 Systems for Human Identity Profiling

The development of systems for human identity testing that integrate all of the required processes has in some ways been an even more challenging goal than the development of many of the sample-to-answer medical diagnostic assays. The reason for this is that medical diagnostic assays typically need to detect only the presence of an amplified DNA sequence from a targeted microorganism. Human identity testing for kinship analysis or forensic purposes, however, presents the need for identifying specific length variants for each of the STR genetic markers that are targeted for analysis. This necessitates the expansion of basic integrated PCR systems to include a size-fractionation process. In conventional forensic laboratories this is performed using CE. Incorporating this into fully integrated systems has proven to be challenging, but significant progress has been made. Successes have been achieved by interfacing nucleic acid extraction and amplification cartridges with standard CE modules or microfluidic CE. Several parallel efforts have resulted in the commercial availability of fully integrated devices that integrate all of the workflow required for sample-to-profile human identity testing.

The RapidHIT™ 200 System (Fig. 5.2) provides an illustration of this more complex workflow (Holland and Wendt, 2015; Hennessy et al., 2013). The RapidHIT™ 200 is a fully integrated sample-to-profile DNA identification system that uses four disposable cartridges in conjunction with an instrument base unit. The sample and control cartridges integrate the nucleic acid extraction and PCR amplification processes. The anode cartridge contains linear



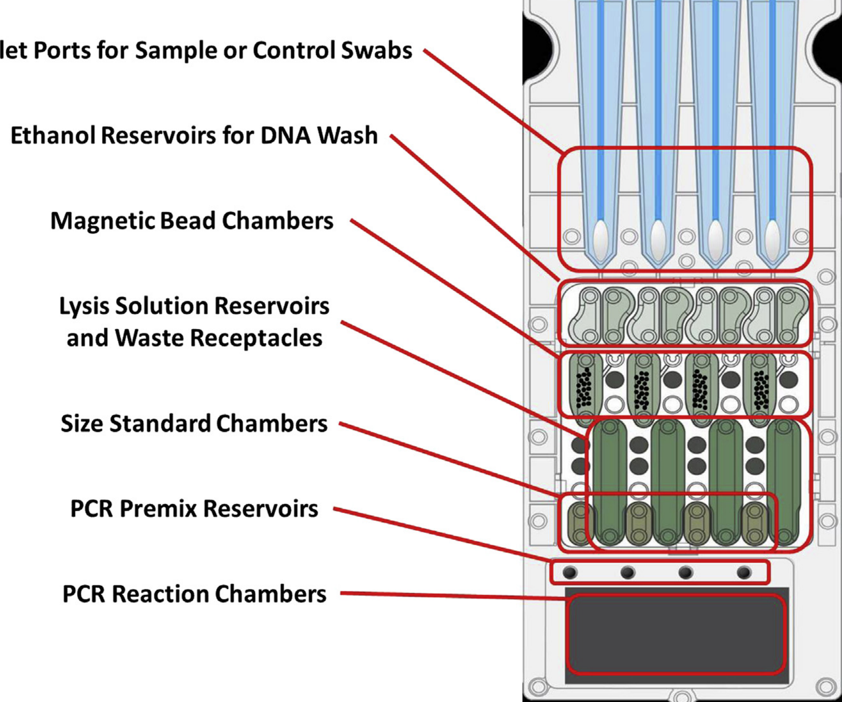


FIGURE 5.2 Photographic image (left) and schematic diagram (right) showing the salient features of a RapidHIT™ 200 assay system sample cartridge. After sample swabs are placed in the inlet ports, a lysis solution is pumped from the lysis solution reservoir to the swab. Suspended cells from the swab are then pumped to the magnetic bead chambers where the cells are lysed, and the DNA is immobilized on magnetic beads. Ethanol is pumped sequentially from each of the ethanol reservoirs to wash the immobilized DNA in the bead chamber. The used ethanol wash is deposited into the waste receptacle. Beads with purified DNA are transferred to the polymerase chain reaction (PCR) chamber and mixed with PCR premix. After PCR amplification, amplicons are fractionated off-cartridge. *Images adapted from IntegenX.*

polyacrylamide for capillary electrophoresis, an agent that dynamically coats the walls of the capillaries to reduce electroosmotic flow during electrophoresis and the anode electrode for electrophoresis. The buffer cartridge serves as the reservoir for both the CE cathode buffer and water for other processes. It is also the receptacle for waste material.

The sample and control cartridges are injection molded to produce an integrated fluidic device consisting of a series of chambers and fluid channels that interface with pneumatically driven valves and pumps that are externally actuated to transport samples and mix fluidic streams (Grover et al., 2003). The sample cartridge has four sample ports and dedicated fluidic pathways to process up to four swabs of cellular material (typically buccal swabs). The control cartridge processes one sample port and dedicated fluidic pathway. The remaining three fluidic pathways are reserved for an internal positive control (a known DNA standard); a negative control (typically no sample) to monitor for contamination of the reagents and/or the fluidic path by extraneous human DNA; and an allelic ladder, which is prepackaged in the size-standard chamber of the flow path and serves as a standard for allele determination during data analysis.

The RapidHIT™ base unit into which the cartridges are inserted for processing has four primary subsystems: one each for sample preparation, fractionation, detection, and control/analysis. The sample preparation subsystem manages the operation of the sample and control cartridges through the use of pneumatics to move liquids from one chamber to the next. Lysis employs a combination of heat and a chaotropic guanidinium salt to release DNA, which is then captured and purified using solid phase extraction on magnetic beads. The beads with bound DNA are then transferred to the reaction chamber and immobilized by fixed magnets. A multiplex PCR master mix with primers for either 16 or 24 STR loci (the PowerPlex® 16HS Fast and GlobalFiler® Express chemistries, respectively) is then added to the reaction chamber and amplified using a Peltier thermocycler. The amplified products are then moved to the size-standard chamber where they are mixed with an internal lane size standard before being pumped to the separation subsystem. After being received by the separation subsystem, which is a separate module housed in the base unit, the linear polyacrylamide from the anode cartridge is used to fill the eight separation capillaries. The amplified samples are then heat denatured, electrokinetically injected, and size fractionated by CE using buffer from the buffer cartridge. After the run, the capillaries are automatically cleaned to prepare them for the next run. The detection subsystem uses a solid-state laser to excite the fluorescently tagged STR products, which are then detected by a charge-coupled device camera as they pass by the capillary detection window. The data on fluorescence intensity and CE mobility are collected and stored by the control/analysis subsystem, which houses the embedded computer that controls all instrument operations. This subsystem automatically processes the raw data for noise filtering, baseline subtraction, spectral deconvolution, and primer peak removal. A local copy of GeneMarker® HID human identity software then identifies the specific genotypes at all STR loci in order to generate a report with the DNA profile of each sample. The total processing time from swab to DNA profile is approximately 90 min.

The DNAscan™ (Fig. 5.3) represents an alternative approach for the generation of STR-based human DNA profiles. It employs a single-use, disposable microfluidic cassette for all DNA profiling processes, including size fractionation of the PCR amplicons, together with an instrument base unit that manages the workflow (Tan et al., 2013). The functional core of the system is an injection-molded BioChipSet Cassette (BCSC), which is constructed from four major components: the smart cartridge, the gel smart cartridge, the integrated biochip, and the separation and detection biochip. The BCSC is preloaded with all of the necessary reagents for sample processing. Liquid reagents are stored in reservoirs that have aluminum foil seals bonded to both ends. Pneumatic pressure is used to burst the seals, thereby releasing the contents of the reservoir. Lyophilized reagents (e.g., PCR master mix, internal lane size standard, and allelic ladders) are also preloaded within the chambers of the BCSC and are reconstituted as the liquid phase reaction products move through the BCSC.

The smart cartridge can process up to five buccal swabs at a time. As with most other systems, the nucleic acid purification method employs chaotropic guanidinium salt lysis followed by guanidinium-mediated binding of DNA to a solid phase silica surface. The DNA extract is then PCR amplified with a PCR master mix containing primers to amplify 16 STR loci (i.e., the PowerPlex® 16 chemistry). An alternative 27-locus multiplex PCR master mix has also been developed and tested. The gel smart cartridge contains the linear polyacrylamide sieving matrix and the electrophoresis buffer, which is used for microfluidic size fractionation and detection. A significant difference from the RapidHIT™ 200 system is that rather than traditional capillaries, the separation and detection biochip performs electrophoresis in six independent microfluidic channels that are 22.5 cm long with a cross-sectional dimension of 40 μm × 100 μm. In order to minimize potential interference from plastic autofluorescence, the entire BCSC is fabricated from a cyclic olefin polymer that produces less autofluorescence than glass. The fluorescent labels on the amplified STR fragments are excited by a 488-nm solid state laser in the optical subsystem of the base unit. Laser light for excitation is transmitted to the detection window of the separation and detection biochip, and the resulting fluorescence is then detected by a series of photomultiplier

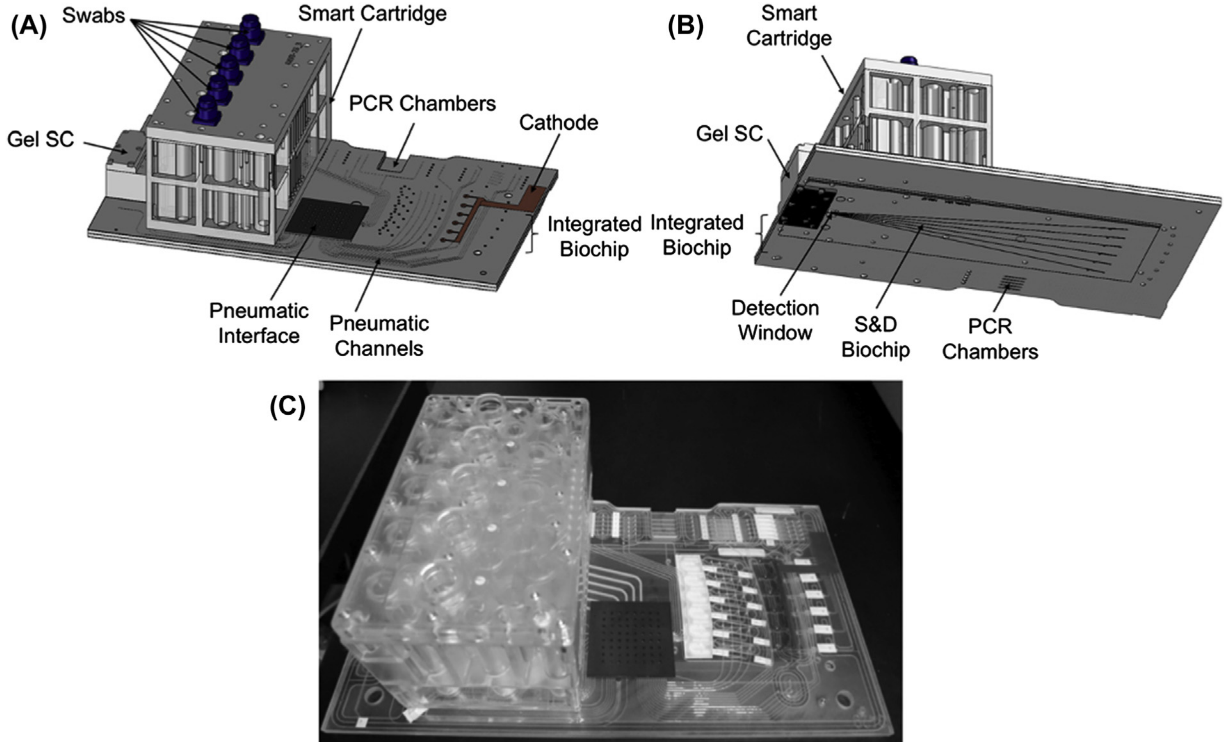


FIGURE 5.3 Structure of the DNAscan™ BioChipSet Cassette (BCSC). (A) Top view schematic showing the location of three of the four components of the BCSC (smart cartridge, gel smart cartridge, and integrated biochip). (B) Bottom view schematic showing the location of the separation and detection biochip, which contains the microfluidic CE channels. (C) Photograph of the BCSC. The primary interfaces with the instrument are via the pneumatic manifold, the polymerase chain reaction chambers, the cathode and anode (not shown, adjacent to the gel smart cartridge), and the separation and detection window. *PCR*, polymerase chain reaction; *gel SC*: gel smart cartridge; *S&D*: separation and detection. *Image from Tan, E., Turingan, R.S., Hogan, C., Vasantgadkar, S., Palombo, L., Schumm, J.W., Selden, R.F., 2013. Fully integrated, fully automated generation of short tandem repeat profiles. Invest. Genet. 4, 1. Creative Commons Attribution License.*

tube detectors. The integrated biochip component contains microfluidic channels and chambers that represent the core of the BCSC. It interfaces with the smart cartridge and facilitates nucleic acid purification by providing the means for liquid transfer from chamber to chamber. The integrated biochip also accepts purified DNA from the smart cartridge in preparation for processing to perform PCR, after which its interface with the separation and detection biochip provides for the transfer of electrophoresis-ready samples to the separation channels.

Upon completion of a sample run, the raw electrophoretic data is processed to achieve baseline subtraction, spectral deconvolution, and to identify signal peaks to produce an interpretable electropherogram. Finally, expert system algorithms are used for automated allele calling. The software interprets the processed data to determine the genotypes for all amplified loci, which collectively form the DNA profile for the analyzed samples. Similar to the RapidHIT™ 200 system, the DNAscan™ system is able to generate a DNA profile through fully automated, fully integrated processing of buccal swabs in just under 90 min.

5.3 CLINICAL APPLICATIONS: PERFORMANCE FOR INFECTIOUS PATHOGEN DIAGNOSTICS

5.3.1 Respiratory Pathogens

The respiratory system is one of the most frequent sites for colonization by infectious agents. Respiratory infections are also an important driver of patient hospitalization. Several sample-to-answer assays are available for the detection of a wide variety of respiratory pathogens. PCR for influenza virus detection has many advantages over conventional diagnostic techniques. For example, detecting influenza in culture takes several days. Antigen-detection-based tests, while simple and fast, suffer from low sensitivity or require the somewhat more tedious examination of cells by fluorescence microscopy. By contrast, nucleic acid tests are fast, sensitive, specific, and well suited to sample-to-answer platforms.

The FilmArray® Respiratory Panel (RP) first gained US Food and Drug Administration (FDA) approval in 2011. The

current iteration of the FilmArray® RP targets the detection of 20 microbes, including numerous viral strains and species of bacteria (Table 5.2). Detection sensitivities vary for each organism and for different versions of the assay (Doern et al., 2013). Overall, the sensitivities reported by independent researchers vary slightly from those reported by the manufacturer in Table 5.2 but generally fall within the 90–100% range for all targeted pathogens, except for Influenza B and adenovirus which have sensitivities of 73% and 83%, respectively (Couturier et al., 2013). While the assay is intended for use with nasopharyngeal swabs, good sensitivities have also been reported for other types of specimens, including nose and throat swabs, sputum, and specimens collected from the lower respiratory tract (Branche et al., 2014; Ruggiero et al., 2014).

The Cepheid Xpert® Flu assay detects and discriminates among Influenza A, Influenza A/H1N1, and Influenza B using either nasopharyngeal swabs or nasal aspirates. It has an overall sensitivity of approximately 95%, though some researchers have seen a somewhat lower sensitivity for Influenza B (Dugas et al., 2014; Li et al., 2012). The Xpert® Flu/RSV XC is a different Xpert® assay cartridge that differentiates Influenza A, Influenza B, and Respiratory Syncytial Virus (RSV). Studies indicate that when compared with other molecular assays, the sensitivity of the Xpert® Flu/RSV XC assay for these viruses ranged from a low of 89% to more than 95% for Influenza A (Popowitch and Miller, 2015; Salez et al., 2013, 2015).

Other sample-to-answer assays approved for respiratory pathogens include the Simplexa™ Flu A/B & RSV assay, the Verigene® Respiratory Viral Plus (RV+) panel, and the Liat® Influenza A and B assay panel. Compared to in-house RT-PCR, the sensitivities of the Simplexa™ A/B & RSV assay for Influenza A, B, and RSV were 97%, 98%, and 99%, respectively (Woodberry et al., 2013). The reported sensitivity and specificity for the Verigene® RV+ was greater than 95% (Cho et al., 2015; Boku et al., 2013). Compared to the Simplexa Flu A/B & RSV assay, the Liat® Influenza A and B assay panel showed greater than 99% sensitivity for Influenza A and 100% for Influenza B (Binnicker et al., 2015).

Mycobacterium tuberculosis is the pathogen responsible for TB, which is a highly contagious and airborne disease that ranks alongside HIV/AIDS as a leading cause of mortality around the globe (WHO, 2014). Drug resistance has become a major obstacle to effective TB treatment and prevention. The spread of TB has been fueled by improper patient treatment, poor management of drug prescriptions and drug quality, and airborne transmissions of bacteria in public places. Strategies to control the spread of TB, including drug-resistant TB, have relied upon slow and labor-intensive diagnostic methods such as the acid-fast bacilli (AFB) smear test (Lawn and Nicol, 2011). Although this has been used for TB detection for more than a century, it has a low specificity and sensitivity (CDC, 2013). Accordingly, the development of faster and more accurate diagnostic tests for *M. tuberculosis* has been a top priority in the medical community (WHO, 2009). A sample-to-result molecular test for TB would have obvious advantages over conventional smear or culture tests.

The Xpert® MTB/RIF assay detects TB and resistance to rifampin in sputum specimens by targeting a TB-specific sequence. Mutations within the *rpoB* gene, which confer resistance to rifampin and are often an early marker of multidrug-resistant TB, are also detected by the assay (Boehme et al., 2010; Helb et al., 2010). Based on results from multiple evaluation studies, the sensitivity of the assay varied from 70–100% in AFB smear test-positive patients. The specificity of the assay ranged from 91–100%. The assay's ability to detect rifampin resistance exceeded 95% sensitivity and specificity (WHO, 2010). Results from controlled clinical validation studies of individuals suspected of having TB showed the sensitivity of a single direct Xpert® MTB/RIF assay to be 73% in smear-negative/culture-positive samples and 90% when three smear-negative samples were tested. The specificity of the Xpert® MTB/RIF assay was 99%. The sensitivity and specificity of rifampin resistance detection were 99% and 100%, respectively (Boehme et al., 2010; WHO, 2010).

This is not to suggest that the Xpert® MTB/RIF assay is without limitations. The need for conventional drug resistance testing is not eliminated, as it is still necessary for the detection of resistance to antibiotics other than rifampin. A negative Xpert® MTB/RIF assay result does not exclude the diagnosis of TB. The Centers for Disease Control and Prevention (CDC) reports that 15–20% of TB cases in the United States that are reported with negative culture results may also have negative nucleic acid amplification test results. This may be due to the low bacterial load or the presence of inhibitory substances in sputum specimens (CDC, 2014).

5.3.2 Gastrointestinal Pathogens

Infectious diarrhea impacts millions of people around the globe each year and is associated with high rates of morbidity and mortality, especially among children. The clinical presentation of infectious gastroenteritis is not informative in terms of specifying an etiologic agent, because diarrhea is the primary symptom caused by a wide range of causative agents. The challenge of selecting an appropriate pathogen identification assay, therefore, is compounded by the fact that for many pathogens, there is no diagnostic test available at all (Buss et al., 2015). This may partially account for the fact that the etiologic agent of infectious diarrhea is only identified in about 20% of patients (Scallan et al., 2011). The rapid and accurate detection of gastrointestinal (GI) pathogens is vital to ensuring

TABLE 5.2 Characteristics of Pathogens Detected by the FilmArray Respiratory panel®

Pathogen	Classification (genome)	Season of highest incidence	Infection demographics	Sensitivity (prospective)	Sensitivity (retrospective)	Specificity (prospective)
Adenovirus	Adenovirus (DNA)	Late winter to early summer	All ages, immunocompromised	88.90%	100%	98.30%
Bocavirus	Parvovirus (DNA)	No peak season	All ages	66.70%	100%	99.80%
Coronavirus HKU1	Coronavirus (RNA)	Winter, spring	Children, adults	95.80%	n/a	99.80%
Coronavirus NL63	Coronavirus (RNA)	Winter, spring	Children, adults	95.80%	n/a	100%
Coronavirus 229E	Coronavirus (RNA)	Winter, spring	Children, adults	100%	n/a	99.80%
Coronavirus OC43	Coronavirus (RNA)	Winter, spring	Children, adults	100%	n/a	99.60%
Human Metapneumovirus	Paramyxovirus (RNA)	Winter, early spring	Children	94.60%	n/a	99.20%
Human Rhinovirus/ Enterovirus	Picornavirus (RNA)	Summer, fall, spring	All ages	92.70%	95.70%	94.60%
Influenza A	Orthomyxovirus (RNA)	Winter	All ages, 5–20% of US population	90.00%	n/a	99.80%
Influenza A/H1	Orthomyxovirus (RNA)	Winter	All ages, 5–20% of US population	n/a	100%	100%
Influenza A/H3	Orthomyxovirus (RNA)	Winter	All ages, 5–20% of US population	n/a	100%	100%
Influenza A/H1-2009	Orthomyxovirus (RNA)	Winter	All ages, 5–20% of US population	88.90%	100%	99.60%

Continued

TABLE 5.2 Characteristics of Pathogens Detected by the FilmArray Respiratory panel®—cont'd

Pathogen	Classification (genome)	Season of highest incidence	Infection demographics	Sensitivity (prospective)	Sensitivity (retrospective)	Specificity (prospective)
Influenza B	Orthomyxovirus (RNA)	Winter	All ages, 5–20% of US population	n/a	100%	100%
Parainfluenza Virus 1	Paramyxovirus (RNA)	Fall, periodicity of 1–2 years	Infants, young children, immunocompromised	100%	97.10%	99.90%
Parainfluenza Virus 2	Paramyxovirus (RNA)	Fall, periodicity of 1–2 years	Infants, young children, immunocompromised	87.40%	100%	99.80%
Parainfluenza Virus 3	Paramyxovirus (RNA)	Spring, summer	Infants, young children, immunocompromised	95.80%	100%	99.80%
Parainfluenza Virus 4	Paramyxovirus (RNA)	Unknown	All ages	100%	100%	99.90%
Respiratory Syncytial Virus	Paramyxovirus (RNA)	Winter, varies by location	Children, older adults	100%	n/a	89.10%
Bordetella pertussis	Bacterium (DNA)	No peak season	All ages	100%[a]	100%	99.90%
Chlamydophila pneumoniae	Bacterium (DNA)	No peak season	Older children, young adults, immunocompromised	100%[a]	n/a	100%
Mycoplasma pneumoniae	Bacterium (DNA)	Summer, periodicity of 4–7 years	Older children, young adults	100%[a]	90.00%	100%

CE, capillary electrophoresis; *NAAT*, nucleic acid amplification test; *RT-PCR*, reverse transcription-polymerase chain reaction.
[a]*Based on less than ten positive samples due to low clinical prevalence in the prospective study.*

appropriate therapy and infection control strategies (Liu et al., 2012; Khare et al., 2014).

Traditionally, bacterial/viral culture, microscopy to search for ova and parasites, and antigen-detection assays were the methods of choice for the identification of GI pathogens. More recently, singleplex real-time PCR assays have been developed to identify specific pathogens. These methods are often time-consuming, labor-intensive, and often offer little more than one shot in the dark after another in an effort to identify a causative pathogen. For the physician, this is not an attractive situation with respect to patient management, infection control, or public health strategies (de Boer et al., 2010). The ability to simultaneously screen for the presence of a wide variety of potential pathogens through multiplex PCR is an important feature of many sample-to-answer assays.

The FilmArray® GI panel and the Verigene® enteric pathogens (EP) panel are two such assays, both of which are commercially available and FDA approved. The FilmArray® GI panel is designed to detect 22 targets (5 viruses, 13 bacteria, and 4 parasites), and the Verigene® EP panel is designed to detect 9 targets (2 viruses, 5 bacteria, and 2 toxins) in a single assay (Khare et al., 2014).

The FilmArray® GI panel showed overall a greater than 90% level of sensitivity and specificity. In a study involving over 1500 cases, the GI panel was reported to have 100% sensitivity/positive predictive value for 12 out of 22 targets (*Plesiomonas shigelloides*, *Salmonella* spp., *Yersinia enterocolitica*, Enterotoxigenic *Escherichia coli*, Shiga toxin-producing *E. coli*, *E. coli* O157, *Cryptosporidium* spp., *Cyclospora cayetanensis*, *Giardia lamblia*, Astrovirus, Rotavirus A, and Sapovirus). For the remaining 10 targets, the sensitivity/positive predictive value was >94.5%. Another advantage of the GI panel is the ability to detect multiple pathogens in a single assay. The use of conventional methods identified more than one pathogen in just 8.3% samples. By contrast, the GI panel indicated the presence of a mixed infection in 31.5% of the specimens (Buss et al., 2015). An assessment of 611 prospective and 839 contrived specimens conducted with the smaller Verigene® EP panel showed sensitivities of 97% for *Salmonella* spp., 93% for *Campylobacter* spp., 100% for *Shigella* spp., 100% for toxin gene *stx*1, and 97% for toxin gene *stx*2 (Novak SM et al., 2014). Taken together, these studies demonstrated that the integrated sample-to-answer diagnostic systems consistently outperform traditional culture-based methods.

Toxigenic *C. difficile* is the quintessential hospital-acquired pathogen in that it is antibiotic-driven and resistant to multiple antibiotics. The estimated prevalence of *C. difficile* infections may be as high as 50% in hospitalized patients where *C. difficile* infection is endemic, 5–7% in residents of long-term care facilities, and generally less than 2% in ambulatory adults. Carriage rates are higher in hospitalized patients who have unrelated conditions that require long-term treatment with antibiotics, which kill off other intestinal bacteria that would normally keep *C. difficile* in check. The reported incidence of *C. difficile* colitis among hospitalized inpatients ranges from 3.8 to 9.5 cases per 10,000 patient days. Rates tend to increase in proportion to the duration of a hospital stay (Dubberke et al., 2008; Kyne et al., 2002; Cohen et al., 2010). *C. difficile* causes a spectrum of diseases, ranging from antibiotic-associated diarrhea to pseudomembranous colitis (Sewell et al., 2014). The fast and accurate detection of *C. difficile* infections is important for appropriate antibiotic treatment and proper control of infection outbreaks.

There are numerous well-established methods available for the detection of *C. difficile* in stool specimens, including stool culture, toxigenic culture, antigen detection, enzyme immunoassay, and molecular testing (Karen and John, 2011). Stool culture is the most sensitive test available, but it is laborious, does not detect toxin production, and has a lengthy turnaround time of up to 4 days. Tissue culture cytotoxicity assays require more technical expertise, but they are slightly faster with a turnaround time of up to 2 days. Still, they are less specific and sensitive than PCR or toxigenic culture-based assays. Antigen-detection assays based on latex agglutination or immunochromatography are faster, producing results in under 1 h. Antigen tests, however, are nonspecific for pathogenic strains. Enzyme immunoassays, which are designed to detect *C. difficile* toxin A, toxin B, or both A and B, are fast and economical but are relatively insensitive.

Here again sample-to-answer diagnostic assays offer superior sensitivity and specificity with a fast turnaround time. The FDA-approved FilmArray® GI panel discussed previously includes a single multiplexed assay (Cdiff) for the identification of toxigenic *C. difficile*. This assay targets both the toxin A and the toxin B genes (*tcdA* and *tcdB*). Common toxigenic strains of *C. difficile* express both toxins, making the presence of either indicative of a pathogenic strain. Empirical testing and in silico sequence analyses indicate that all toxinotypes will be detected by the assay, including the epidemic North American Profile 1 NAP1/027/BI hypervirulent strain, which has been identified as a cause of hospital outbreaks worldwide (Chapin et al., 2011), Although they are detected, these strains are not specifically differentiated by the GI assay.

The Xpert® *C. difficile* and Xpert® *C. difficile*/Epi assays have both been approved by the FDA and are performed directly from a stool sample. The former targets the *C. difficile tcdB* and has been shown to have a sensitivity of 94% and a specificity of 96% (Novak-Weekley et al., 2010). The latter assay differentiates the hypervirulent NAP1/027/BI strain. Similarly, the Verigene® *C. difficile* assay has the ability to detect both toxin-encoding *tcdA* and *tcdB* genes, and it identifies the NAP1/027/BI strain

(Carroll et al., 2013). Regardless of the assay used, rapid sample-to-answer molecular testing for *C. difficile* makes it possible to facilitate a reduction in transmission during outbreaks by providing cost-effective and timely detection of the pathogen (Sewell et al., 2014).

Another GI pathogen of concern is norovirus, the most common cause of acute gastroenteritis, which is highly contagious via the fecal-oral route. Transmission of the disease is associated with crowded living environments such as cruise ships, school dormitories, daycare centers, and prisons. Accordingly, prompt identification of norovirus early in an outbreak can have a positive impact by informing appropriate actions to prevent further spread of the disease.

The Cepheid Xpert® Norovirus assay is similar in design to other GeneXpert® assays. It is performed directly from stool specimens and distinguishes between Norovirus genogroups I and II. A study of over 1400 samples found that compared to a CDC composite reference method, the Xpert® Norovirus assay demonstrated a positive percent agreement for genogroups I and II of 98% and 99%, respectively, and a negative percent agreement of 98% for both genogroups I and II (Gonzalez et al., 2016).

5.3.3 Sexually Transmitted Pathogens

Prompt diagnosis and treatment of urinary tract infections (UTIs) and sexually transmitted infections (STIs) can be critical to a positive patient outcome. Complications from delays in treatment or ineffective treatment due to misdiagnoses can result in an increased risk of acquiring other pathogens, such as HIV; dissemination of an existing infection to the circulatory and lymphatic systems; the development of pelvic inflammatory disease, which can lead to infertility; and complications with pregnancy, including damage to the fetus and/or miscarriage. The ability to rapidly diagnose such infections, particularly those that may be asymptomatic but still transmissible, enhances disease management and public health by improving the rate of treatment and the timely notification of sexual partners for follow-up testing.

Traditional diagnostic assay methods for the causative agents of UTIs and STIs include direct culture, serology, and immunochromatographic tests and nucleic acid amplification-based tests. Of these, the DNA amplification assays are generally preferred for the detection of *Chlamydia trachomatis*, *Neisseria gonorrhoeae*, and *Trichomonas vaginalis* (Papp et al., 2014). For these pathogens, the lengthy turnaround time from specimen collection to diagnosis results in a delayed administration of appropriate therapy and a decreased efficacy of infection control strategies (Liu et al., 2012; Khare et al., 2014).

As potentially useful as sample-to-answer assay systems would be in this sphere of infectious disease management, there is a narrower variety of FDA-approved or clinical laboratory improvement amendments-waived tests available. For example, there are no approved sample-to-result molecular assays for syphilis. New assays for the diagnosis and quantitation of HIV are still in the development stage. There is no FDA-approved FilmArray® panel for UTI or STI pathogens, but a first-generation STI panel was designed to detect and identify nine common STI pathogens (*C. trachomatis*, *N. gonorrhoeae*, *Treponema pallidum*, *T. vaginalis*, *Mycoplasma genitalium*, *Haemophilus ducreyi*, herpes simplex virus 1 and 2, *Ureaplasma urealyticum*). This panel was used to test 295 clinical specimens from 190 subjects. The STI panel results were compared to results from standard clinical tests performed on duplicate specimens. These included gram staining, wet mount examination, viral culture, and the serum syphilis IgG test. Concordance between the FilmArray® STI panel and standard testing was 83% for *T. vaginalis*, 98% for *C. trachomatis*, and 97% for *N. gonorrhoeae*. The assay also detected *T. pallidum* in samples from four patients who were subsequently diagnosed with syphilis by serology.

The Cepheid Xpert® CT/NG assay detects *C. trachomatis* and *N. gonorrhoeae* on vaginal swabs, endocervical swabs, and male and female urine. The sensitivity and specificity for both targets were found to exceed 95% and 99%, respectively (Causer et al., 2014). Illustrating the potential impact of rapid diagnostic sample-to-answer platforms on public health in a point-of-care context, asymptomatic men were tested on-site at a sexual health clinic using the Xpert® CT/NG assay. The rapid turnaround time of the assay enabled infected patients (14% of those tested) to receive appropriate treatment 2 days after their test. In contrast to this, the turnaround time for specimens that were tested at an off-site laboratory using conventional methods was 10 days (Gaydos, 2014). One area for potential improvement in existing assays is the lack of validation studies to support the testing of rectal and pharyngeal specimens. This would be useful, as many gonococcal and chlamydial infections of the rectal and oropharyngeal tissues are asymptomatic. Testing of 409 rectal swabs showed the sensitivity and specificity of the Xpert® CT/NG assay to be 86% and 99.2% for *C. trachomatis* and 91.1% and 100% for *N. gonorrhoeae*, respectively (Goldenberg et al., 2012).

Infection by herpes simplex virus (HSV)-1 and HSV-2 is a common cause of genital and oral STIs. After an initial acute phase infection, the viruses typically enter a latent phase. While the rate of subclinical HSV shedding tends to decrease after the first year of infection following the initial clinical episode, viral shedding may persist at high rates in some infected individuals for several years after infection (Phipps et al., 2011). In addition, HSV reactivation from latency after the primary infection can cause a clinical recurrence of the local disease accompanied by high rates of viral shedding. The continued risk that this poses both

for HSV horizontal transmission to sexual partners and vertical transmission from mother to infant makes the diagnosis of patients who may be actively shedding important to disease management. The Simplexa™ HSV 1 & 2 Direct Kit was approved for genital swab samples in 2015. According to the manufacturer, the sensitivity and specificity of the assay from genital swabs for HSV 1 and HSV 2 exceeds 97% (Focus Diagnostice, 2015).

5.3.4 Central Nervous System Pathogens

Infections of the central nervous system (CNS) are notable for their diversity and the unique challenges they present due to the potential morbidity and mortality that they cause in conjunction with inherent difficulties involved in their diagnosis and treatment. Patients with CNS infections may display depressed levels of consciousness, photophobia, altered mental states, fever, lethargy, and a wide range of other symptomology. The etiologic agents of CNS infections may range from viruses and bacteria to fungi and even parasites. Such infections often produce changes in the cerebrospinal fluid (CSF), which makes it a valuable specimen for diagnostic analyses.

Viral infection is the most common form of aseptic meningitis, and enteroviruses are the most common viral cause, particularly in pediatric cases (Hong et al., 2015). While the disease is generally self-limiting and is typically treated with supportive therapy, it can be difficult based on symptomology to differentiate it from early stage bacterial meningitis, which is a far more serious disease that can lead to death within hours or leave patients with permanent brain damage, hearing loss, and learning disabilities. Because of the potential seriousness of CNS infections, it is not unusual for a patient to be admitted and treated with broad-spectrum antibiotics until a clear diagnosis is made. Therefore it would be reasonable to postulate that the rapid and accurate diagnosis of enteroviral-associated meningitis infections would help to prevent the unnecessary use of antibiotics, shorten the duration of hospitalization, and reduce healthcare costs (Ramers et al., 2000; King et al., 2007). This represents an ideal context for accurate and fast sample-to-answer assay systems.

The FilmArray® Meningitis/Encephalitis Panel was FDA approved in 2015. The panel targets 14 pathogens for detection (*E. coli* K1, *Haemophilus influenzae*, *Listeria monocytogenes*, *Neisseria meningitidis*, *Streptococcus agalactiae*, *Streptococcus pneumoniae*, *cytomegalovirus*, enterovirus, HSV-1, HSV-2, HHV-6, human parechovirus, varicella zoster virus, and *Cryptococcus neoformans*). A preclinical assessment of the panel using CSF specimens reported positive and negative agreements across methods of 93% and 92%, respectively. Results obtained with the Meningitis/Encephalitis Panel were compared to results with routine testing methods, and discrepancies were resolved through the use of additional nucleic acid amplification tests or by direct sequencing (Hanson et al., 2015). A second study using the panel with an HIV-infected population in Uganda detected *Cryptococcus* in the CSF of patients diagnosed with a first episode of cryptococcal meningitis by fungal culture with 100% sensitivity and specificity (Rhein et al., 2016).

The Cepheid Xpert® EV assay also tests for the presence of enterovirus RNA in CSF. The manufacturer reports that the specificity and sensitivity of the assay exceed 96% and 97%, respectively. An independent set of studies compared patients with aseptic meningitis who had been diagnosed using the Xpert® EV assay to patients diagnosed on the basis of a conventional in-house PCR assay and to patients who had not been diagnosed at all. Those patients for whom a diagnosis was confirmed on the Xpert® EV had a significantly shorter duration of broad-spectrum antibiotic administration, and fewer patients received acyclovir. The average length of stay for these patients was only 0.5 days, compared with 2 days and 4 days for patients in the group with conventional PCR or no diagnosis, respectively. The reported sensitivity for this assay was 95−100%, and the specificity was 100% (Giulieri et al., 2015; Marlowe et al., 2008; Kost et al., 2007).

In addition to their disease-causing potential in STIs, HSV-1 and HSV-2 can also cause encephalitis, with HSV-1 being more common in pediatric cases. Immediate treatment with the antiviral drug acyclovir is indicated for those patients suspected (based on symptomology) of having HSV-associated encephalitis. Left untreated, the mortality rate for these patients approaches 70% (Raschilas et al., 2002). Ironically, the administration of powerful antiviral drugs to a patient may actually interfere with traditional diagnosis by viral culture, resulting in false negative results. Such adverse effects do not impact PCR-based assays, which makes the detection of HSV using either the FilmArray® Meningitis/Encephalitis Panel or the Simplexa™ HSV 1 & 2 Direct Kit (which is approved for use with CSF specimens) faster and more reliable diagnostic options.

5.4 FORENSIC APPLICATIONS: PERFORMANCE FOR HUMAN IDENTITY TESTING

Around the world, there has been a rapid growth in demand by law enforcement agencies for Human DNA Identity testing (i.e., DNA profiling) in connection with criminal investigations. This has driven the development of more expedient techniques to handle the increasing number of samples being submitted for analysis. Multiarray capillary electrophoretic instruments that allow for the simultaneous analysis of multiple samples as well as direct PCR

amplification methodologies, which eliminate the need for extraction and quantification, have both accelerated individual steps of the DNA analysis process. Still, accurate human DNA profiling of trace and otherwise challenging samples still requires substantial hands-on time by trained forensic analysts in centralized laboratories where segmented workflows are used to extract DNA, amplify target genetic markers, fractionate amplicons, and interpret data to produce a meaningful DNA profile.

The full integration of all steps in the DNA processing workflow into a compact system has been termed "rapid DNA." Automated rapid DNA systems allow for the generation of full STR DNA profiles in a fraction of the time required by conventional laboratory methods with minimal user intervention. This provides forensic investigators and law enforcement with the ability to identify possible perpetrators more quickly, which represents an advantage when trying to obtain a warrant or to apprehend a suspect. Aside from assisting law enforcement and military intelligence with the generation of investigative leads, these rapid DNA systems also provide forensic laboratories with faster sample processing times for the analysis of reference samples.

The full potential value of these sample-to-profile systems became evident after the US Supreme Court decision in Maryland v. King, 133S.Ct. 1958 (2013). The court ruled that the collection and analysis of a buccal swab from an arrestee for DNA profiling purposes was a legitimate police booking procedure. Moreover, the court ruled that the arrestee's DNA profile could be used by law enforcement to search a criminal DNA database as part of determining whether or not to release the individual who was arrested. This decision made instantly obvious the need for and value of being able to generate DNA profiles of arrestees at police booking stations rather than at off-site forensic laboratories.

The first fully integrated rapid DNA system for human identification was the RapidHIT™ 200 Human DNA Identification System from IntegenX. When used with the GlobalFiler® Express assay, the system produces a DNA profile consisting of 21 autosomal and 3 sex-determining markers from buccal swabs and other sample types in less than 2 h. Regardless of its speed, this integrated system would not be of use for forensic purposes if it did not meet acceptable sensitivity, precision, and accuracy standards. When used with reference quality buccal swabs, the platform has over an 88% success rate in producing complete profiles (Jovanovich et al., 2015). When a dilution series of DNA is placed onto swabs and introduced into the platform, full profiles are obtained with 200 ng of DNA, and partial profiles are obtained with as little as 10 ng of input DNA. Buccal swabs analyzed on the RapidHIT™ 200 system have been found to be 100% concordant with profiles generated by traditional laboratory methods, and resultant alleles size within 0.5 bp of corresponding alleles in the allelic ladder, demonstrating acceptable precision (Hennessy et al., 2014). These data support the use of the RapidHIT™ 200 system for the analysis of single-source buccal samples for the expedient profiling of reference-quality samples.

Aside from aiding forensic laboratories in processing reference samples more quickly, there are multiple applications for rapid DNA testing of reference samples. Federal immigration officials are investigating the use of this technology for the analysis of reference type samples to verify that children entering the United States are related to their accompanying adults. The Department of Homeland Security would like to employ rapid DNA testing to support or reject claims of familial relatedness that are used to justify permission to immigrate. At borders and ports, this technology could help to ensure that individuals entering the county are not in terrorist DNA databases. Another rapid DNA system, the DNAscan™ from Healthcare and NetBio, was the first rapid DNA system approved for the upload of generated reference DNA profiles into the National DNA Index System by the FBI. This clearly demonstrates that these systems have the ability to generate actionable intelligence for law enforcement (Tan et al., 2013).

In the United States, police agencies in Arizona, Florida, and South Carolina, to name a few, have also begun using this platform for the generation of investigative leads for casework samples. The rate of DNA profiling success for casework samples appears to be sample-dependent, but full profiles have been generated from cigarette butts (range of success 0−100%, n = 29 samples from 6 donors); drinking items (range of success 6−100%, n = 13 samples from 10 donors); and chewing gum (range of success 0−100%, n = 23 samples from 16 donors) (Verheij et al., 2013). Given this high variability in success rates for profile detection with forensic type samples, additional improvements in the sensitivity of these systems will be needed to allow for their application to the analysis of more challenging sample types, such as touch or contact DNA samples, which typically have only trace quantities of DNA.

When DNA quantities are not as limiting as with touch type samples, it has been demonstrated that it is possible to produce full DNA profiles using traditional laboratory methodologies following a reextraction of samples previously analyzed by the RapidHIT™ system. This indicated that the RapidHIT™ system does not consume all available biological material (Thong et al., 2015; Verheij et al., 2013). This is important considering the limited nature of many forensic type samples, since it demonstrates that if further efficiencies in DNA extraction can be achieved, the success rate with casework samples might be improved.

Police are interested in the analysis of casework samples such sample-to-profile integrated systems in order to

determine, prior to suspect release, whether an individual is potentially connected to an item of evidence from a crime scene. Military applications for casework samples include analyses of munitions and weapons to determine attribution. Currently, however, FBI policy requires all casework samples to undergo a human-specific quantification step. As this step is not part of the integrated workflow on any sample-to-profile system, casework samples analyzed using rapid DNA systems must still be reanalyzed following a traditional laboratory workflow. This exemplifies how advances in technology can push ahead of existing policy. The more widespread use of sample-to-profile systems for casework samples, therefore, will necessitate a change in official policy, taking into account the technical capabilities of these new platforms. Alternatively, a quantification module would need to be added to rapid DNA systems, even though it is not necessary for the accurate and reliable operation of the instrument.

5.5 CONTINUING EVOLUTION OF SAMPLE-TO-ANSWER TECHNOLOGIES

The development of PCR technology initiated a transformational change in the field of molecular biology. The rapid growth in the "big data analytics" of the human genome has been a major driving force behind a similar transformational change in medical diagnostics. Where health care once meant diagnostic and treatment options tailored to the physiology of the average patient, physicians and other healthcare professionals now speak of the promise of "Precision Medicine" personalized to a patient's unique physiology and genetic background. An important part of making personalized care a reality is the ability to obtain patient test data in a manner that is timelier than that possible using traditional labor-intensive assay methods. One of the important goals of molecular diagnostics, therefore, has always been the development of faster and more cost-effective approaches to performing diagnostic assays. Initially, this took the form of an emphasis on large-scale automation and high-throughput instrumentation in centralized laboratories using one-size-fits-all segmented workflows.

By bringing analyses closer to the patient in the form of "Point-of-Care Diagnostics" it was hoped that several process steps could be eliminated, thereby facilitating a shorter time to result, a faster health management response, better therapeutic turnaround times, and ultimately a greater opportunity for improved patient outcomes. Making this goal a reality, however, was not to be achieved by relying on large-scale automation but rather on process integration and miniaturization driven largely by advances in engineering guided by the diagnostic needs of physicians and patients. Specifically, advances in microfluidics and microscale automation made it possible to fully integrate, on a handheld scale, the previously segmented processes of nucleic acid extraction and purification, PCR amplification, amplicon detection, and even such complex tasks as electrophoretic size fractionation. With miniaturization also came the ability to reduce processing times so as produce answers more quickly. So simple, reliable, and user friendly are these fully integrated diagnostic systems that the terms "sample-to-answer" and "sample-to-profile" testing have entered the common lexicon of both the molecular diagnostic and the forensic/human identity testing communities.

The economic forces of the diagnostics industry have motivated both well-established players like GE and a host of new market entrants to develop and make commercially available a rapidly expanding selection of diagnostic and DNA profiling tools. A major area of focus of these commercial systems has been assays for infectious disease diagnostics. Given that this industry is still in its infancy, however, there is still enormous room for growth and technological improvement. The menu of available assays can be expanded to new pathogen panels, drug resistance panels, and assays for genetic markers of human disease predisposition. Cepheid has already moved in this direction with their Xpert® FII & FV assay for human gene variants associated with thrombophilia. Of course as infectious pathogens mutate, there will be a need for modified assays, and existing assays can be improved for the use of more patient sample types. On the engineering side, continued advances in material science and fabrication technologies will drive the development of true point-of-care diagnostics instruments that are smaller, faster, and portable enough to be used at a patient's bedside.

While this industry is well positioned for a bright future, one potential obstacle that will need to be addressed in regard to sample-to-answer devices is cost, not so much for the instrument base units but rather the costs per assay, which, in most cases, are appreciably higher than the price of more traditional assay methods. Of course it can be argued that, as with all new technologies, costs will invariably drop as the technology matures. Moreover, since many of these assays target multiple pathogens in a single assay run, the cost per pathogen targeted provides users with a better overall value than if the same series of tests had all been performed on an à la carte basis. It also may be worth asking if the ability to easy test for a wide range of pathogens in a simple 1-h test will produce a shift in how physicians think about patient diagnostics. Will multiplex panels reduce the amount of time that physicians would normally spend trying to narrow a preliminary diagnosis to a "most likely causative agent" that can be tested for? Will routine testing for pathogens that might not rise to the top of a physician's list of "most likely suspects" provide new insights on the complexity of disease processes and ensure more responsive care for those patients whose conditions do not fall within the most probable etiology?

The importance of validation studies demonstrating the reliability, precision, and accuracy of sample-to-answer assays cannot be overstated, but these will need to go hand-in-hand with rigorous assessments of the impact of these systems on patient care. It is necessary but not sufficient that an assay be "fit for purpose". Ultimately, the long-term success of these amazing systems will rest on the ability to show a clear value. For sample-to-profile systems used for human identity testing, that value might be quantified in terms of crimes prevented or solved. For medical diagnostic assays, it will be necessary to demonstrate that their use results in shorter hospital stays, the elimination of unnecessary or ineffectual treatments, improved patient outcomes, and an overall quantifiable reduction in healthcare costs.

REFERENCES

Binnicker, M.J., Espy, M.J., Irish, C.L., Vetter, E.A., 2015. Direct detection of influenza A and B viruses in less than 20 minutes using a commercially available rapid PCR assay. J. Clin. Microbiol. 53, 2353–2354.

Boehme, C.C., Nabeta, P., Hillemann, D., Nicol, M.P., Shenai, S., Krapp, F., Allen, J., Tahirli, R., Blakemore, R., Rustomjee, R., 2010. Rapid molecular detection of tuberculosis and rifampin resistance. N. Engl. J. Med. 363, 1005–1015.

Boku, S., Naito, T., Murai, K., Tanei, M., Inui, A., Nisimura, H., Isonuma, H., Takahashi, H., Kikuchi, K., 2013. Near point-of-care administration by the attending physician of the rapid influenza antigen detection immunochromatography test and the fully automated respiratory virus nucleic acid test: contribution to patient management. Diagnostic Microbiol. Infect. Dis. 76, 445–449.

Branche, A.R., Walsh, E.E., Formica, M.A., Falsey, A.R., 2014. Detection of respiratory viruses in sputum from adults by use of automated multiplex PCR. J. Clin. Microbiol. 52, 3590–3596.

Buchan, B.W., Ledeboer, N.A., 2014. Emerging technologies for the clinical microbiology laboratory. Clin. Microbiol. Rev. 27, 783–822.

Buss, S.N., Leber, A., Chapin, K., Fey, P.D., Bankowski, M.J., Jones, M.K., Rogatcheva, M., Kanack, K.J., Bourzac, K.M., 2015. Multicenter evaluation of the BioFire FilmArray gastrointestinal panel for etiologic diagnosis of infectious gastroenteritis. J. Clin. Microbiol. 53, 915–925.

Butt, S., Maceira, V., Mccallen, M., Stellrecht, K., 2014. Comparison of three commercial RT-PCR systems for the detection of respiratory viruses. J. Clin. Virol. 61, 406–410.

Carroll, K.C., Buchan, B.W., Tan, S., Stamper, P.D., Riebe, K.M., Pancholi, P., Kelly, C., Rao, A., Fader, R., Cavagnolo, R., 2013. Multicenter evaluation of the Verigene *Clostridium difficile* nucleic acid assay. J. Clin. Microbiol. 51, 4120–4125.

Causer, L.M., Hengel, B., Natoli, L., Tangey, A., Badman, S.G., Tabrizi, S.N., Whiley, D., Ward, J., Kaldor, J.M., Guy, R.J., 2014. A field evaluation of a new molecular-based point-of-care test for chlamydia and gonorrhoea in remote Aboriginal health services in Australia. Sex. Health 12, 27–33.

CDC, 2013. Diagnosis of Tuberculosis Disease. Centers for Disease Control and Prevention, Core Curriculum on Tuberculosis: What the Clinician Should Know. Centers for Disease Control and Prevention.

CDC, 2014. Report of an Expert Consultation on the Uses of Nucleic Acid Amplification Tests for the Diagnosis of Tuberculosis.

Chapin, K.C., Dickenson, R.A., Wu, F., Andrea, S.B., 2011. Comparison of five assays for detection of Clostridium difficile toxin. J. Mol. Diagn. 13, 395–400.

Chen, L., Tian, Y., Chen, S., Liesenfeld, O., 2015. Performance of the Cobas® Influenza A/B Assay for Rapid Pcr-Based Detection of Influenza Compared to Prodesse ProFlu^{+} and Viral Culture. Eur. J. Microbiol. Immunol. (Bp) 5, 236–245.

Cho, H.J., Jang, J.W., Ko, S.Y., Choi, S.H., Lim, C.S., An, S.S.A., 2015. Evaluation and verification of the nanosphere Verigene RV^{+} assay for detection of influenza A/B and H1/H3 subtyping. J. Med. Virol. 87, 18–24.

Cohen, S.H., Gerding, D.N., Johnson, S., Kelly, C.P., Loo, V.G., Mcdonald, L.C., Pepin, J., Wilcox, M.H., 2010. Clinical practice guidelines for *Clostridium difficile* infection in adults: 2010 update by the society for healthcare epidemiology of America (SHEA) and the infectious diseases society of America (IDSA). Infect. Control Hosp. Epidemiol. 31, 431–455.

Couturier, M.R., Barney, T., Alger, G., Hymas, W.C., Stevenson, J.B., Hillyard, D., Daly, J.A., 2013. Evaluation of the FilmArray® Respiratory Panel for clinical use in a large children's hospital. J. Clin. Lab. Anal. 27, 148–154.

De Boer, R.F., Ott, A., Kesztyüs, B., Kooistra-Smid, A.M., 2010. Improved detection of five major gastrointestinal pathogens by use of a molecular screening approach. J. Clin. Microbiol. 48, 4140–4146.

Doern, C.D., Lacey, D., Huang, R., Haag, C., 2013. Evaluation and implementation of FilmArray version 1.7 for improved detection of adenovirus respiratory tract infection. J. Clin. Microbiol. 51, 4036–4039.

Dubberke, E.R., Butler, A.M., Reske, K.A., Agniel, D., Olsen, M.A., D'angelo, G., Mcdonald, C., Fraser, V.J., 2008. Attributable outcomes of endemic *Clostridium difficile*-associated disease in nonsurgical patients. Emerg. Infect. Dis. 4.

Dugas, A.F., Valsamakis, A., Gaydos, C.A., Forman, M., Hardick, J., Kidambi, P., Amin, S., Gupta, A., Rothman, R.E., 2014. Evaluation of the Xpert Flu rapid PCR assay in high-risk emergency department patients. J. Clin. Microbiol. 52, 4353–4355.

Focus Diagnostics, 2015. In: Diagnostics, F. (Ed.), Simplexa HSV 1 & 2 direct. Focus Diagnostics, Cypress, CA.

Fujii, K., Inokuchi, S., Kitayama, T., Nakahara, H., Mizuno, N., Sekiguchi, K., 2013. A comparison of DNA extraction using AutoMate Express™ and EZ1 advanced XL from liquid blood, bloodstains, and semen stains. J. Forensic Sci. 58, 981–988.

Gaydos, C.A., 2014. Review of use of a new rapid real-time PCR, the Cepheid GeneXpert®(Xpert) CT/NG assay, for *Chlamydia trachomatis* and *Neisseria gonorrhoeae*: results for patients while in a clinical setting. Expert Rev. Mol. Diagn. 14, 135–137.

Giljohann, D.A., Seferos, D.S., Daniel, W.L., Massich, M.D., Patel, P.C., Mirkin, C.A., 2010. Gold nanoparticles for biology and medicine. Angew. Chem. Int. Ed. Engl. 49, 3280–3294.

Giulieri, S.G., Chapuis-Taillard, C., Manuel, O., Hugli, O., Pinget, C., Wasserfallen, J.-B., Sahli, R., Jaton, K., Marchetti, O., Meylan, P., 2015. Rapid detection of enterovirus in cerebrospinal fluid by a fully-automated PCR assay is associated with improved management of aseptic meningitis in adult patients. J. Clin. Virol. 62, 58–62.

Goldenberg, S., Finn, J., Sedudzi, E., White, J., Tong, C., 2012. Performance of the GeneXpert CT/NG assay compared to that of the Aptima AC2 assay for detection of rectal *Chlamydia trachomatis* and

Neisseria gonorrhoeae by use of residual Aptima Samples. J. Clin. Microbiol. 50, 3867−3869.

Gonzalez, M.D., Langley, L.C., Buchan, B.W., Faron, M.L., Maier, M., Templeton, K., Walker, K., Popowitch, E.B., Miller, M.B., Rao, A., 2016. Multicenter evaluation of the Xpert Norovirus assay for detection of Norovirus genogroups I and II in fecal specimens. J. Clin. Microbiol. 54, 142−147.

Grover, W.H., Skelley, A.M., Liu, C.N., Lagally, E.T., Mathies, R.A., 2003. Monolithic membrane valves and diaphragm pumps for practical large-scale integration into glass microfluidic devices. Sensors Actuators B Chem. 89, 315−323.

Haned, H., Gill, P., 2015. Expert Systems in DNA Interpretation. Wiley Encyclopedia of Forensic Science.

Hanson, K., Slechta, E., Killpack, J., Heyrend, C., Lunt, T., Daly, J., Hemmert, A., Blaschke, A., 2016. Preclinical assessment of a fully automated multiplex PCR panel for the detection of central nervous system pathogens. J. Clin. Microbiol. JCM 54, 785−787.

Helb, D., Jones, M., Story, E., Boehme, C., Wallace, E., Ho, K., Kop, J., Owens, M.R., Rodgers, R., Banada, P., 2010. Rapid detection of *Mycobacterium tuberculosis* and rifampin resistance by use of on-demand, near-patient technology. J. Clin. Microbiol. 48, 229−237.

Hennessy, L.K., Franklin, H., Li, Y., Buscaino, J., Chear, K., Gass, J., Mehendale, N., Williams, S., Jovanovich, S., Harris, D., 2013. Developmental validation studies on the RapidHIT™ human DNA identification system. Forensic Sci. Int. Genet. Suppl. Ser. 4, e7−e8.

Hennessy, L.K., Mehendale, N., Chear, K., Jovanovich, S., Williams, S., Park, C., Gangano, S., 2014. Developmental validation of the GlobalFiler® express kit, a 24-marker STR assay, on the RapidHIT® System. Forensic Sci. Int. Genet. 13, 247−258.

Holland, M., Wendt, F., 2015. Evaluation of the RapidHIT™ 200, an automated human identification system for STR analysis of single source samples. Forensic Sci. Int. Genet. 14, 76−85.

Hong, J., Kim, A., Hwang, S., Cheon, D.-S., Kim, J.-H., Lee, J.-W., Park, J.-H., Kang, B., 2015. Comparison of the genexpert enterovirus assay (GXEA) with real-time one step RT-PCR for the detection of enteroviral RNA in the cerebrospinal fluid of patients with meningitis. Virol. J. 12, 1.

Jovanovich, S., Bogdan, G., Belcinski, R., Buscaino, J., Burgi, D., Butts, E.L., Chear, K., Ciopyk, B., Eberhart, D., El-Sissi, O., 2015. Developmental validation of a fully integrated sample-to-profile rapid human identification system for processing single-source reference buccal samples. Forensic Sci. Int. Genet. 16, 181−194.

Karen, C.C., John, G.B., 2011. Biology of *Clostridium difficile*: implications for epidemiology and diagnosis. Annu. Rev. Microbiol. 65, 501−521.

Khare, R., Espy, M.J., Cebelinski, E., Boxrud, D., Sloan, L.M., Cunningham, S.A., Pritt, B.S., Patel, R., Binnicker, M.J., 2014. Comparative evaluation of two commercial multiplex panels for detection of gastrointestinal pathogens by use of clinical stool specimens. J. Clin. Microbiol. 52, 3667−3673.

King, R.L., Lorch, S.A., Cohen, D.M., Hodinka, R.L., Cohn, K.A., Shah, S.S., 2007. Routine cerebrospinal fluid enterovirus polymerase chain reaction testing reduces hospitalization and antibiotic use for infants 90 days of age or younger. Pediatrics 120, 489−496.

Kost, C.B., Rogers, B., Oberste, M.S., Robinson, C., Eaves, B.L., Leos, K., Danielson, S., Satya, M., Weir, F., Nolte, F.S., 2007. Multicenter beta trial of the GeneXpert enterovirus assay. J. Clin. Microbiol. 45, 1081−1086.

Kyne, L., Hamel, M.B., Polavaram, R., Kelly, C.P., 2002. Health care costs and mortality associated with nosocomial diarrhea due to *Clostridium difficile*. Clin. Infect. Dis. 34, 346−353.

Lawn, S.D., Nicol, M.P., 2011. Xpert® MTB/RIF assay: development, evaluation and implementation of a new rapid molecular diagnostic for tuberculosis and rifampicin resistance. Future Microbiol. 6, 1067−1082.

Lawn, S.D., Mwaba, P., Bates, M., Piatek, A., Alexander, H., Marais, B.J., Cuevas, L.E., Mchugh, T.D., Zijenah, L., Kapata, N., 2013. Advances in tuberculosis diagnostics: the Xpert MTB/RIF assay and future prospects for a point-of-care test. Lancet Infect. Dis. 13, 349−361.

Li, M., Brenwald, N., Bonigal, S., Chana, K., Osman, H., Oppenheim, B., 2012. Rapid diagnosis of influenza: an evaluation of two commercially available RT-PCR assays. J. Infect. 65, 60−63.

Liu, L., Johnson, H.L., Cousens, S., Perin, J., Scott, S., Lawn, J.E., Rudan, I., Campbell, H., Cibulskis, R., Li, M., 2012. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. Lancet 379, 2151−2161.

Marlowe, E.M., Novak, S.M., Dunn, J.J., Smith, A., Cumpio, J., Makalintal, E., Barnes, D., Burchette, R.J., 2008. Performance of the GeneXpert® enterovirus assay for detection of enteroviral RNA in cerebrospinal fluid. J. Clin. Virol. 43, 110−113.

Miao, B., Peng, N., Li, L., Li, Z., Hu, F., Zhang, Z., Wang, C., 2015. Centrifugal microfluidic system for nucleic acid amplification and detection. Sensors 15, 27954−27968.

Novak, S.M., Bobenchik, A., Cumpio, J., Marlowe, E., 2014. Evaluation of the Verigene EP IUO test for the rapid detection of bacterial and viral causes of gastrointestinal infection, abstr 1317. Abstr. 30th Clin. Virol. Symp. Annu. Meet. Pan. Am. Soc. Clin. Virol.

Novak-Weekley, S.M., Marlowe, E.M., Miller, J.M., Cumpio, J., Nomura, J.H., Vance, P.H., Weissfeld, A., 2010. *Clostridium difficile* testing in the clinical laboratory by use of multiple testing algorithms. J. Clin. Microbiol. 48, 889−893.

Papp, J.R., Schachter, J., Gaydos, C.A., Van Der Pol, B., 2014. Recommendations for the laboratory-based detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae*—2014. MMWR. Recommendations and reports: morbidity and mortality weekly report. Recomm. Reports Centers Dis. Control 63, 1.

Park, S., Zhang, Y., Lin, S., Wang, T.-H., Yang, S., 2011. Advances in microfluidic PCR for point-of-care infectious disease diagnostics. Biotechnol. Adv. 29, 830−839.

Phipps, W., Saracino, M., Magaret, A., Selke, S., Remington, M., Huang, M.-L., Warren, T., Casper, C., Corey, L., Wald, A., 2011. Persistent genital herpes simplex virus-2 shedding years following the first clinical episode. J. Infect. Dis. 203, 180−187.

Popowitch, E.B., Miller, M.B., 2015. Performance characteristics of Xpert flu/RSV XC assay. J. Clin. Microbiol. 53, 2720−2721.

Poritz, M.A., Blaschke, A.J., Byington, C.L., Meyers, L., Nilsson, K., Jones, D.E., Thatcher, S.A., Robbins, T., Lingenfelter, B., Amiott, E., 2011. FilmArray, an automated nested multiplex PCR system for multi-pathogen detection: development and application to respiratory tract infection. PloS One 6, e26047.

Ramers, C., Billman, G., Hartin, M., Ho, S., Sawyer, M.H., 2000. Impact of a diagnostic cerebrospinal fluid enterovirus polymerase chain reaction test on patient management. JAMA 283, 2680−2685.

Raschilas, F., Wolff, M., Delatour, F., Chaffaut, C., De Broucker, T., Chevret, S., Lebon, P., Canton, P., Rozenberg, F., 2002. Outcome of and prognostic factors for herpes simplex encephalitis in adult patients: results of a multicenter study. Clin. Infect. Dis. 35, 254−260.

Rhein, J., Bahr, N.C., Hemmert, A.C., Cloud, J.L., Bellamkonda, S., Oswald, C., Lo, E., Nabeta, H., Kiggundu, R., Akampurira, A., 2016. Diagnostic performance of a multiplex PCR assay for meningitis in an HIV-infected population in Uganda. Diagn. Microbiol. Infect. Dis. 84, 268−273.

Ruggiero, P., Mcmillen, T., Tang, Y.-W., Babady, N.E., 2014. Evaluation of the BioFire FilmArray respiratory panel and the GenMark eSensor respiratory viral panel on lower respiratory tract specimens. J. Clin. Microbiol. 52, 288−290.

Sackmann, E.K., Fulton, A.L., Beebe, D.J., 2014. The present and future role of microfluidics in biomedical research. Nature 507, 181−189.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., Erlich, H.A., 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239, 487−491.

Salez, N., De Lamballerie, X., Zandotti, C., Gazin, C., Charrel, R., 2013. Improved sensitivity of the novel Xpert flu test for detection of influenza B virus. J. Clin. Microbiol. 51, 4277−4278.

Salez, N., Nougairede, A., Ninove, L., Zandotti, C., De Lamballerie, X., Charrel, R.N., 2015. Prospective and retrospective evaluation of the Cepheid Xpert® Flu/RSV XC assay for rapid detection of influenza A, influenza B, and respiratory syncytial virus. Diagn. Microbiol. Infect. Dis. 81, 256−258.

Scallan, E., Hoekstra, R.M., Angulo, F.J., Tauxe, R.V., Widdowson, M.-A., Roy, S.L., Jones, J.L., Griffin, P.M., 2011. Foodborne illness acquired in the United States—major pathogens. Emerg. Infect. Dis. 17.

Sewell, B., Rees, E., Thomas, I., Ch'ng, C.L., Isaac, M., Berry, N., 2014. Cost and impact on patient length of stay of rapid molecular testing for *Clostridium difficile*. Infect. Dis. Ther. 3, 281−293.

Strohmeier, O., Keller, M., Schwemmer, F., Zehnle, S., Mark, D., Von Stetten, F., Zengerle, R., Paust, N., 2015. Centrifugal microfluidic platforms: advanced unit operations and applications. Chem. Soc. Rev. 44, 6187−6229.

Tan, E., Turingan, R.S., Hogan, C., Vasantgadkar, S., Palombo, L., Schumm, J.W., Selden, R.F., 2013. Fully integrated, fully automated generation of short tandem repeat profiles. Invest. Genet. 4, 1.

Thong, Z., Phua, Y.H., Loo, E.S., Goh, S.K., Ang, J., Looi, W.F., Syn, C.K.C., 2015. Evaluation of the RapidHIT™ 200 system: a comparative study of its performance with Maxwell® DNA IQ™/Identifiler® Plus/ABI 3500xL workflow. Forensic Sci. Int. Genet. 19, 22−27.

Tortoli, E., Russo, C., Piersimoni, C., Mazzola, E., Dal Monte, P., Pascarella, M., Borroni, E., Mondo, A., Piana, F., Scarparo, C., 2012. Clinical validation of Xpert MTB/RIF for the diagnosis of extrapulmonary tuberculosis. Eur. Respir. J. 40, 442−447.

Verheij, S., Clarisse, L., Van Den Berge, M., Sijen, T., 2013. RapidHIT™ 200, a promising system for rapid DNA analysis. Forensic Sci. Int. Genet. Suppl. Ser. 4, e254−e255.

WHO, 2009. Global Tuberculosis Control 2009. World Health Organization.

WHO, 2010. Roadmap for Rolling Out Xpert MTB/RIF for Rapid Diagnosis of TB and MDR-TB. World Health Organization, Geneva, pp. 1−12.

WHO, 2014. World Health Organization Global Tuberculosis Report 2013.

Woodberry, M.W., Shankar, R., Cent, A., Jerome, K.R., Kuypers, J., 2013. Comparison of the Simplexa FluA/B & RSV direct assay and laboratory-developed real-time PCR assays for detection of respiratory virus. J. Clin. Microbiol. 51, 3883−3885.

Xu, M., Qin, X., Astion, M.L., Rutledge, J.C., Simpson, J., Jerome, K.R., Englund, J.A., Zerr, D.M., Migita, R.T., Rich, S., 2013. Implementation of FilmArray respiratory viral panel in a core laboratory improves testing turnaround time and patient care. Am. J. Clin. Pathol. 139, 118−123.

Chapter 6

High-Resolution Melting Curve Analysis for Molecular Diagnostics

J.S. Farrar[1] **and C.T. Wittwer**[2]
[1]Virginia Commonwealth University School of Medicine, Richmond, VA, United States; [2]University of Utah School of Medicine, Salt Lake City, UT, United States

6.1 INTRODUCTION TO MELTING ANALYSIS

DNA melting or denaturation occurs when double-stranded DNA (dsDNA) separates into random coils of single-stranded DNA. The melting of DNA can be followed with dyes that fluoresce in the presence of dsDNA. When a polymerase chain reaction (PCR) product is melted in the presence of a dsDNA dye, the fluorescence is monitored continuously and plotted against the temperature. As the temperature increases, there is a characteristic drop in fluorescence at high temperature that coincides with the denaturation of the PCR product. The melting profile of a PCR product depends on its guanine-cytosine (GC) content, length, and sequence. Short PCR products usually melt in a single transition. The melting temperature is the temperature at which 50% of the DNA is double-stranded and the remainder is single-stranded. Longer PCR products often melt in multiple transitions, corresponding to melting domains of different stability. It has become common practice to melt PCR products after amplification to assess their identity and purity.

Traditional genotyping by melting analysis relies on the use of labeled probes. Variants under the probe are detected, but not those outside the probe region. Conventional variant scanning identifies variants anywhere within a PCR product, but it requires separation of the mixture through a gel or other matrix. In contrast, high-resolution melting is a closed-tube, homogeneous technique that does not require labeled probes or sample processing after PCR. Both genotyping and scanning are performed with only a dsDNA dye and two primers. For fine discrimination of multiple variants within a particular region, an unlabeled probe or snapback primer tail can be included. High-resolution melting can rapidly establish sequence identity when specific genotyping is not required, as in human leukocyte antigen (HLA) matching or the assessment of repeated DNA segments. Melting is nondestructive, allowing subsequent analysis of the PCR product when necessary. The simplicity and speed of high-resolution melting analysis are attractive and are reflected by its increasing use in molecular diagnostics. The method is enabled by high-resolution melting instruments and dsDNA dyes that detect heteroduplexes. High-resolution melting analysis has previously been reviewed (Reed et al., 2007; Erali et al., 2008a; Montgomery et al., 2010; Vossen et al., 2009). Additional reviews that focus on specific applications include microbiology (Ruskova and Raclavsky, 2011; Tong and Giffard, 2012), genetics (Er and Chang, 2012), scanning (Taylor, 2009; Erali and Wittwer, 2010), and single-nucleotide polymorphisms (Bruzzone and Steer, 2015).

6.1.1 History

Fluorescent melting curve analysis was introduced as an integral part of real-time PCR (see Chapter 4) with the LightCycler® in 1997 (Wittwer et al., 1997a). Instead of looking at product fluorescence once every cycle, continuous monitoring during PCR was performed (Wittwer et al., 1997b). The dsDNA dye SYBR® Green I allowed both template quantification and melting curve analysis to distinguish among PCR products (Ririe et al., 1997). In many cases the need for further analysis by electrophoresis was eliminated.

High-resolution amplicon melting was first reported using a 5′-labeled primer to provide the fluorescence signal.

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00006-7
Copyright © 2017 Elsevier Ltd. All rights reserved.

Both heterozygous and homozygous single-base variants were identified (Gundry et al., 2003). However, a fluorescently labeled primer was needed and only variants residing in the melting domain of the labeled primer were detected. The introduction of dsDNA dyes that efficiently detect heteroduplexes eliminated the need for labeled primers (Wittwer et al., 2003). For the first time, heterozygous variants anywhere between the primers could be easily detected with no processing or separation steps after PCR.

The first report of genotyping by fluorescent melting curve analysis used hybridization probes and fluorescence resonance energy transfer (Lay and Wittwer, 1997). A fragment of *F5* containing the variant g.1691 A > G (Leiden) was amplified by asymmetric PCR. The reaction included a Cy5 labeled primer and a 3′-fluorescein labeled probe covering the variant site. When the fluorescein probe was hybridized to the extension product of the Cy5 primer, energy transfer enhanced the Cy5 fluorescence. Melting of the duplexes provides the genotype because the different alleles resulted in different probe melting temperatures. Combined with rapid cycle PCR (Wittwer et al., 1994), amplification and genotyping required only 30 min. Subsequently, genotyping with two adjacent hybridization probes (HybProbes®), each labeled with a donor or acceptor fluorophore, was first demonstrated using *HFE* variants (Bernard et al., 1998). The method was later simplified by using one fluorescein-labeled probe (SimpleProbe®) instead of two probes (Crockett and Wittwer, 2001).

Genotyping by melting without labeled probes was first achieved using SYBR® Green I and allele-specific PCR using three primers, one with a GC tail to discriminate alleles (Germer and Higuchi, 1999). Later, the need for allele-specific PCR and GC tailing was eliminated by small amplicon melting, a method enabled by high-resolution melting. Only two standard primers and a dsDNA dye that detects heteroduplexes were required (Liew et al., 2004). For probe-based genotyping, the same dsDNA dye can detect the melting of unlabeled probes (Zhou et al., 2004a,b). Unlabeled probe genotyping uses three standard oligonucleotides (the probe is blocked at the 3′-end to prevent extension) and can easily identify homozygous variants that may be difficult to genotype by amplicon melting alone.

Alternatively, snapback primer genotyping uses only two primers, one with a 5′-tail that serves as the probe element, forming an intramolecular hairpin (Zhou et al., 2008). Asymmetric PCR is usually required for sufficient single-stranded product to hybridize to unlabeled probes or snapback primers. Two melting regions are observed, one for the probed region and another for the amplicon, allowing simultaneous genotyping and variant scanning in the same assay (Zhou et al., 2005).

In this chapter, we consider only methods that do not require labeled oligonucleotides, focusing on high-resolution melting with dyes that increase in fluorescence when bound to the surface and/or intercalate into dsDNA. Multiplexing by colored probes is common in real-time PCR despite its complexity and cost. Multiplexing by melting temperature (Tm) allows low-cost generic dyes to replace more expensive covalent labels. Higher-order multiplexing by combining Tm and color for two-dimensional analysis can be powerful (Chakravorty et al., 2010), but is not considered here. Similarly, melting analysis of bound duplexes is not covered (Meuzelaar et al., 2007; Marcy et al., 2008). The high-resolution melting methods described here are best used to analyze a limited number of targets in many samples, in contrast to many targets on a few samples where massively parallel sequencing and array methods are supreme. We focus on what any laboratory can do with dyes and simple oligonucleotides in a closed tube. As it turns out, simple, low-cost melting can do much more than most people realize. A common abbreviation for high-resolution melting (HRM™) has been used as a trademark and is not used here. If abbreviations are necessary, "HRMA" is generic and does not have commercial implications.

High-resolution melting has been compared with other analysis methods. It is more sensitive than denaturing high-pressure liquid chromatography (Sumer Celebi and Ozdag, 2014; Tsai et al., 2011; Wang et al., 2015a). When blood cultures are analyzed for bacteria (Jeng et al., 2012) or yeast (Duyvejonck et al., 2015), the concordance between high-resolution melting and mass spectroscopy is high. Most discrepancies resolved in favor of melting, but polymicrobial samples were better detected with mass spectroscopy. Sanger sequencing identifies the position and base change of any variant, but high-resolution melting has fivefold better sensitivity (Gorniak et al., 2015). Massively parallel sequencing has similar sensitivity and produces massive amounts of sequence, but high-resolution melting is faster, less expensive, and easier to use for specific targets (Hinrichs et al., 2015). In studies where high-resolution melting is compared with exonuclease probes and allele-specific PCR, concordance is high and all three assays perform well (Perera et al., 2015; Zhang et al., 2015a,b) with rare genotyping errors in some exonuclease assays (Janukonyte et al., 2010).

6.1.2 Components of the Technology

High-resolution melting of DNA for genotyping, scanning, and sequence matching was made possible by three developments in the early 2000s. First, dsDNA dyes that detect heteroduplexes were identified from existing commercial dyes or by new design and synthesis. Second, dedicated high-resolution melting instruments became

available and existing real-time PCR instruments were modified to increase their melting resolution. Finally, software for data normalization, curve shape comparison, and genotype clustering were developed. Used in concert, these tools increase melting curve quality and allow the detection of small sequence differences in PCR products.

6.1.2.1 Double-Stranded DNA Dyes for Melting Analysis

dsDNA binding dyes monitor the melting of entire PCR products. Traditionally the dye SYBR® Green I was used as a sensitive and convenient dye for quantitative PCR and product melting analysis (Wittwer et al., 1997a). Early reports demonstrated single-base genotyping in products up to 167 base pairs (bp) (Lipsky et al., 2001). However, the protocol required purifying the samples after PCR, followed by the addition of high concentrations of dye. Unfortunately, SYBR Green I inhibits PCR at these high saturating concentrations (Wittwer et al., 1997b). At dye concentrations that are compatible with PCR, lower melting products are difficult to observe, possibly because of dye or strand redistribution during melting and/or GC specificity (Wittwer et al., 2003; Giglio et al., 2003). That is, when multiple products are present after amplification, although the lower Tm products are present, they may not be observed during melting. This includes the lower Tm heteroduplexes formed during amplification of a heterozygote. Indeed, most dsDNA dyes, including SYBR Green I, ethidium bromide, SYBR Gold, Pico Green, TOTO®-1, and YOYO®-1 do not detect heteroduplexes well (Wittwer et al., 2003). Nevertheless, SYBR Green I can detect heterozygotes in a closed-tube system (Dufresne et al., 2006). That is, you do not have to detect heteroduplexes to detect a heterozygote. When heteroduplexes are not detected, low-temperature transitions are not observed and the heterozygote appears between the two homozygote melting curves. SYBR Green I genotyping is consistently successful when the melting differences between genotypes are large, as after gap PCR (Pornprasert et al., 2008), repeat typing (Nguyen et al., 2012), or methylation analysis (Worm et al., 2001). When the Tm difference between homozygotes is small, using a dye that detects heterduplexes will facilitate detection of heterozygotes.

Different dsDNA dyes detect heteroduplexes to a greater or lesser extent. Fig. 6.1 compares LCGreen® Plus with SYBR Green I and two other commercial dyes that have been used to detect heteroduplexes. The apparent heteroduplex percentages observed were 23.5% for LCGreen Plus, 17.6% for SYTO® 9 (Thermo Fisher Scientific), and 15.2% for EvaGreen® (Biotium). The percentage of observed heteroduplexes correlates with the ease of heterozygote detection and scanning sensitivity. SYTO 13 and SYTO 16 may also be suitable for real-time PCR and high-resolution melting (Radvanszky et al., 2015). Unlike SYBR Green I, "saturating" dsDNA dyes do not inhibit PCR at saturating concentrations.

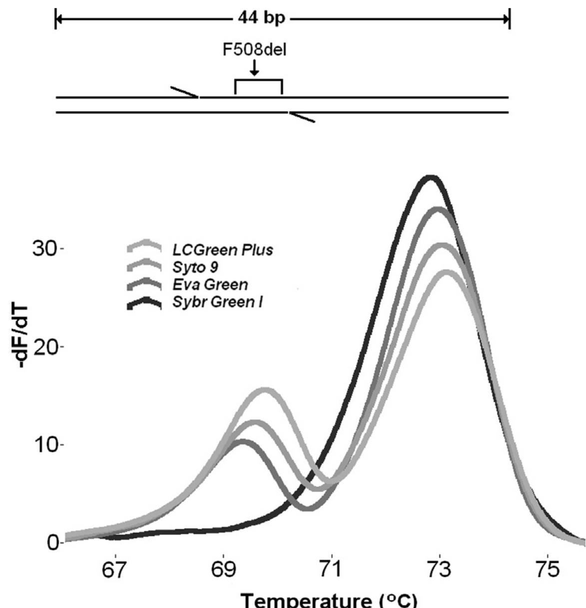


FIGURE 6.1 Heterozygote detection with different dyes. The ability to detect heterozygotes was compared among three saturating dyes (LCGreen® Plus, SYTO® 9, and EvaGreen®), and the nonsaturating dye, SYBR® Green I. A short amplicon surrounding the p.F508del site of *CFTR* was amplified using heterozygous p.F508del DNA as the template. High-resolution melting analysis (HR-1™, BioFire Defense) was performed and the data were plotted as a negative derivative after background removal, normalization, and curve overlay. In each case, optimal dye concentrations were used that minimized polymerase chain reaction inhibition and maximized heteroduplex detection. Although heteroduplexes are not observed with SYBR Green I, increasing heteroduplex signals are observed for EvaGreen, SYTO 9, and LCGreen Plus. *bp*, base pair.

LCGreen dyes are asymmetric cyanines specifically developed for heteroduplex detection (Dujols et al., 2006). SYTO 9 was initially marketed as a nucleic acid stain for both gram-positive and gram-negative bacteria. EvaGreen was developed for DNA quantification in capillary gel electrophoresis and quantitative PCR. dsDNA dyes generally increase DNA melting temperatures and may require adjustment of cycling parameters. If the melting temperature of an amplicon is already high, an agent such as dimethyl sulfoxide (DMSO), glycerol, or betaine may be required to prevent apparent "inhibition." DMSO has been reported to increase scanning sensitivity (Song et al., 2015).

6.1.2.2 Instrumentation

High-resolution melting needs to discriminate small melting curve differences. As PCR products become larger, the differences between them become smaller (Gundry et al., 2003; Reed and Wittwer, 2004) and instrument precision becomes more critical. Standard real-time

instruments are not designed for high-resolution melting. High-resolution melting and real-time PCR platforms have been evaluated using identical melting analysis software (Herrmann et al., 2006, 2007a,b; Li et al., 2014). In the first set of studies (Herrmann et al., 2006, 2007a,b), two technical metrics were evaluated. The first, variation in curve shape, is most important in variant scanning and heterozygote identification. This is the variation remaining after the curves are overlaid to best compare curve shapes and is determined in part by data acquisition density. The second technical metric, variation in absolute temperature, is important for distinguishing different homozygotes that differ by melting temperature. Air-controlled and single-sample systems performed better than did plate-based systems by this metric. These early studies showed that instruments designed for high-resolution melting have lower temperature variation and are better scanning and genotyping platforms.

High-resolution melting assays can often resolve variants even when their Tm difference is predicted to be at or near zero. This is because nearest-neighbor theory does not fully encompass all DNA stability effects. The abilities of high-resolution melting instruments to genotype a nearest-neighbor symmetric, Class 4 single nucleotide variant with a predicted change in melting temperature of 0.00°C (Li et al., 2014) were compared. Using manufacturer-recommended melting conditions, heterozygote detection and homozygote differentiation were assessed in four PCR product lengths of 51–547 bp with LCGreen Plus dye. Over all instruments, heterozygote genotyping accuracy was excellent (99.7%; n = 2141), whereas homozygote accuracy was only 70.3% (n = 4441). Fig. 6.2 shows the variability in homozygote genotyping accuracy for the instruments that were available. Instruments with single-sample detection (individual melting or optical scanning) performed better than did full-plate imaging with temperature variation roughly half that of full-plate instruments. Variations in fluorescent intensity and temperature across a plate contribute to the increased temperature variation. Resolution may also differ among different models of the same manufacturer (Li et al., 2014; Ebili and Ilyas, 2015).

Unfortunately, many manufacturers improved melting resolution simply by increasing the melting time to collect more points, so that melting on some instruments now takes over an hour to perform. In contrast, many of the early, dedicated melting instruments required only 1–5 min for melting curve acquisition with excellent resolution, although some of these are no longer commercially available. Ten data points per degrees centigrade appear adequate to capture the features important in melting curves. The fastest rate reported for high-resolution melting is 0.5°C/s, generating melting curves in <1 min. This has been referred to as "high-speed melting" (Cao et al., 2015; Sundberg et al., 2014). More important than the data density is the accuracy of the temperature measurement.

As shown in Fig. 6.2, accurate temperature measurement and control are not trivial tasks and are further complicated when multiple samples are measured. The mismatch between instrument and solution temperatures during thermal cycling and melting has been studied (Sanford and Wittwer, 2013). Using a temperature-sensitive dye for temperature measurement, the solution temperature lagged behind instrument temperature by up to 8°C during cycling, often requiring 5–10 s at target temperatures for

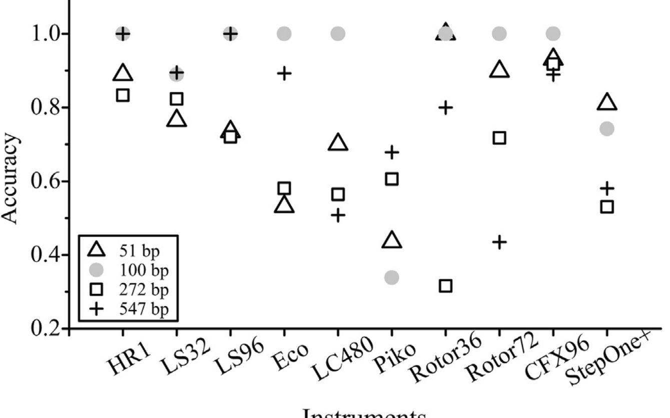


FIGURE 6.2 Homozygous genotyping accuracy on 10 instruments of four polymerase chain reaction product lengths. The instrument melting rates decrease from left to right. Manufacturer-recommended melting conditions were used and melting curve analysis was performed with custom software. Genotyping accuracy was best when single samples were interrogated by either individual melting (HR-1™) or optical scanning of samples (CFX96™). Instrument abbreviations and their manufacturers are as follows: HR-1™ (HR1, BioFire Defense), LightScanner® 32 (LS32, BioFire Defense), LightScanner® 96 (LS96, BioFire Defense), Eco™ (Eco, Illumina), LightCycler® 480 (LC480, Roche), PikoReal™ 96 (Piko, ThermoFisher Scientific), Rotor-Gene® Q (Rotor36/72, 36 and 72 sample carousels, Qiagen), CFX96™ (CFX96, Bio-Rad), and StepOnePlus™ (StepOne+, ThermoFisher Scientific).

equilibrium. Melting curves were also displaced 0.2−1.1°C. Measurement of solution temperature with fluorescence improves temperature accuracy and thus high-resolution melting results. Innovative methods for improving temperature precision are needed as new instruments continue to be introduced. It is hoped that competition among manufacturers will improve melting hardware in the future, because genotyping accuracy and scanning sensitivity depend solely on instrument precision to resolve small differences between melting curves. In addition to the technical performance of melting instruments, there are many other factors to consider. The "best" instrument depends on the primary application and the relative needs for throughput, turn-around time, sample format, and economy.

6.1.2.3 Data Analysis

Melting data are usually processed before final presentation. However, viewing the original data is important for quality control and to establish the data density (points per degrees centigrade). The actual data points (not smoothed curves) should be displayed after analog to digital conversion but before software manipulation. The absolute sample fluorescence depends on the sample volume, dye concentration, instrument optics, and PCR amplification. Samples that amplify poorly (including any negative controls) will have low fluorescence and should not be normalized or further analyzed. If the melting transition is unexpectedly meager or multiple transitions are observed when only one is expected, unintended ("nonspecific") PCR products are likely. It is best first to optimize the PCR by annealing temperature gradients and gels before melting analysis is attempted. Tms and profiles can be predicted (Dwight et al., 2011; Rasmussen et al., 2007) and if the experimental curves are different from these predictions, the PCR is again suspect. Different samples will usually vary in absolute fluorescence as a result of optical and/or volume differences (Fig. 6.3A).

Data analysis software is variable across instruments, with certain features offered by some manufacturers and excluded by others. Some instruments require laborious calibration before high-resolution melting can be performed. Value-added (i.e., expensive) software packages or even monthly software fees may be required. Nevertheless, a few simple steps can turn data that appears highly variable (Fig. 6.3A) into discrete clusters that correlate with genotype (Fig. 6.3C−E).

6.1.2.3.1 Background Removal

High-resolution melting curves are easier to compare after the fluorescence background is removed. Although the sharpest drop in fluorescence occurs as the amplicon melts, fluorescence also decreases with temperature outside the melting transition. This decrease is nearly linear on each side of the derivative peak of the PCR product (Gundry et al., 2003). When not removed, it also elevates the low-temperature baseline on derivative plots. Most commercial software does not remove this background, limiting any match to predicted melting curves because the fluorescence includes both background and duplex melting components. To remove the fluorescence background, linear extrapolation can be used and is the accepted method for absorbance melting curves (Mergny and Lacroix, 2003). This simple method transfers well to fluorescent amplicon melting if only one melting transition is present. However, fluorescence melting introduces new artifacts not seen with absorbance, particularly at low temperatures where transient associations between primers may interact with dye at high concentrations. This background is better approximated by an exponential if: (1) multiple melting transitions are present, or (2) there are both low (probe) and high (amplicon) temperature transitions (Erali et al., 2008b). If the background is properly removed, the fluorescence both before and after duplex melting will be horizontal, correctly converting fluorescence to DNA helicity.

6.1.2.3.2 Normalization

Multiple samples will differ in absolute fluorescence intensity because of optical efficiency differences between samples (both on excitation and emission). Normalization plots all curves between 100% (completely hybridized) and 0% (completely single-stranded). Although differences in fluorescence magnitude are lost with normalization, the melting transitions are much easier to compare. Melting curves after background removal and normalization are shown in Fig. 6.3B.

6.1.2.3.3 Curve Overlay

Overlay of melting curves provides the best sensitivity for heterozygote detection. This is true irrespective of the instrument because all instruments show some temperature variation between samples and all samples are not identical because of pipetting and evaporation variation. These differences can be decreased by overlaying all curves so that they are superimposed over the high-temperature region (usually between 2% and 5% normalized fluorescence). This curve overlay allows for easy identification of heterozygotes by their melting curve shapes. However, absolute temperature differences between samples are lost, so that variant homozygotes are more difficult to identify after curve overlay. In general, when it is important to identify homozygous variants, data should be examined both with and without curve overlay. Overlaid melting curves after background subtraction and normalization are shown in Fig. 6.3C.

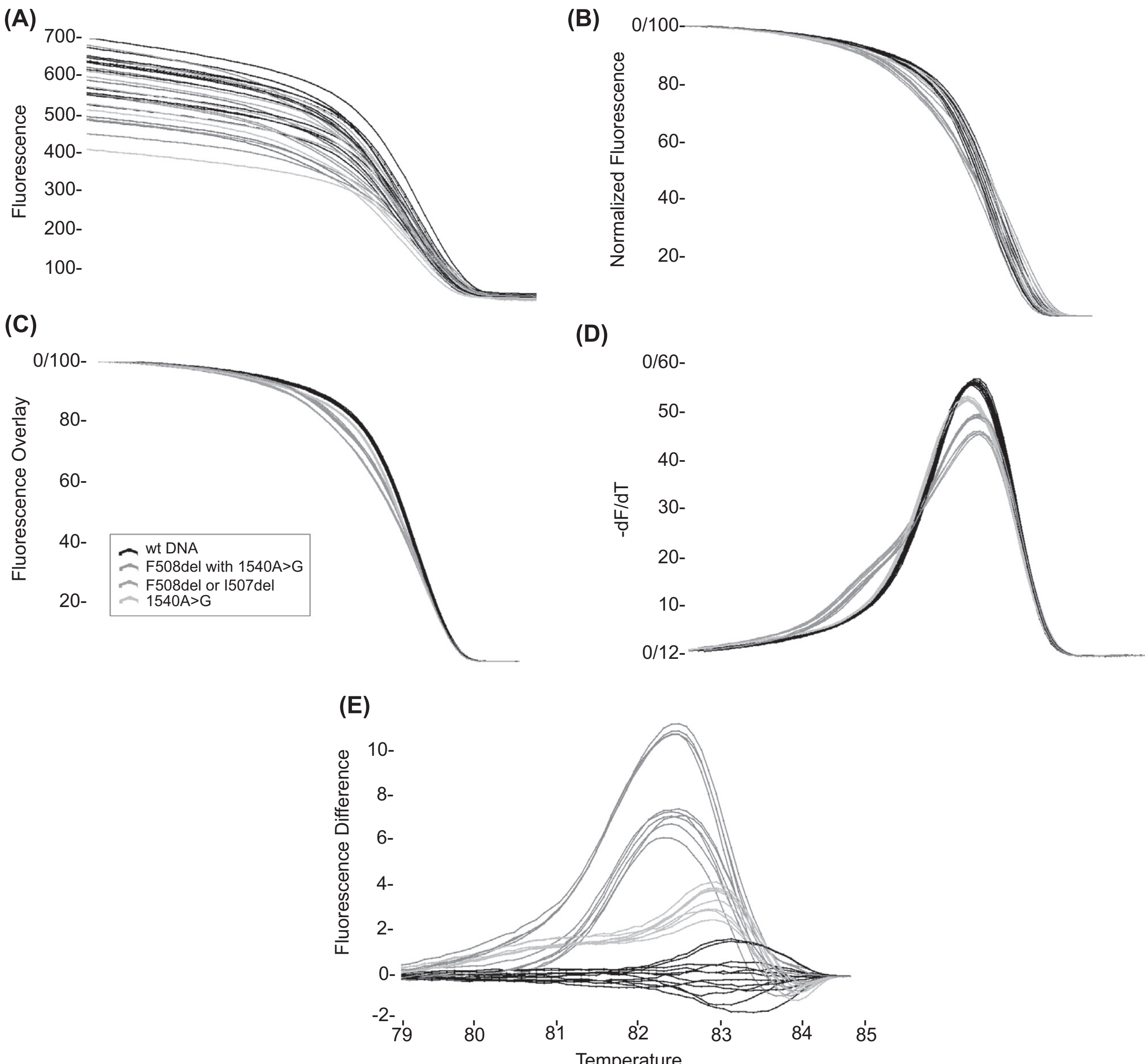


FIGURE 6.3 High-resolution melting curve analysis. Polymerase chain reaction amplicon melting curves reveal duplex melting as a sharp drop in fluorescence with increasing temperature. The data displayed in all panels include four melting curve clusters of *CFTR*, exon 11. (A) The original data, showing variation in fluorescence among samples at low temperature owing to optical or sample differences, do not separate into four clusters. (B) Normalized data, correcting for the temperature effect on fluorescence, still do not completely separate the four clusters. (C) Curve overlay of normalized data compares the shape of the melting curves. Although the differences between clusters are small, they are now discernable. (D) The negative derivative of overlaid data further visualizes the differences between clusters. (E) The preferred data display for high-resolution amplicon melting analysis is shown here as a difference plot in which all clusters are easily discerned.

6.1.2.3.4 Derivative Plots

Negative first derivative plots are used extensively in labeled probe genotyping (Lay and Wittwer, 1997; Bernard et al., 1998) and low-resolution amplicon melting analysis (Ririe et al., 1997). The negative first derivative is usually determined by Savitzky−Golay polynomial estimation (Wittwer and Kusukawa, 2011). When melting curves are converted to derivative plots, curve peaks approximate Tms unless the peaks are asymmetric. Unlabeled probe (Zhou et al., 2004a) and snapback primer genotyping (Zhou et al., 2008) typically use derivative plots for analysis. Although derivative plots can be used for high-resolution amplicon melting (Fig. 6.3D), high-resolution data are best presented without the smoothing process inherent in taking a derivative.

6.1.2.3.5 Difference Plots

High-resolution scanning data are best presented on difference plots (Fig. 6.3E). Different genotypes are easiest to

identify on these magnified graphs (Wittwer et al., 2003). They are created by plotting the difference between a reference curve and all other melting curves. The reference curve is usually the average of all wild-type curves so that this group clusters around the horizontal axis. Different genotypes trace different paths for easy visual discrimination. Automatic clustering of genotypes can be performed, for example, by unbiased hierarchal clustering or trained learning (Zhou et al., 2005; Reja et al., 2010).

6.2 GENOTYPING OF KNOWN VARIANTS BY HIGH-RESOLUTION MELTING

High-resolution melting is a simple yet powerful, closed-tube, homogeneous technique for genotyping. Common methods include small amplicon melting, unlabeled probe, and snapback primer genotyping.

6.2.1 Genotyping by Melting Small Polymerase Chain Reaction Products

Small amplicon melting was the first genotyping method introduced with dsDNA saturation dyes (Wittwer et al., 2003). Although long PCR products (>500 bp) can be genotyped by high-resolution amplicon melting, melting curve differences between genotypes are easier to distinguish in small amplicons (Liew et al., 2004). Very rapid or "extreme" PCR is possible with small amplicons by reducing the denaturation temperature and eliminating all temperature holds (Farrar and Witter, 2014). No probes are required because the dye labels the entire PCR product. Genotyping of a single nucleotide variant is illustrated in Fig. 6.4, which shows normalized and background subtracted melting curves. No curve overlay is performed so that homozygotes are readily identified by a shift in Tm and heterozygotes are easily identified by a change in curve shape.

Biallelic single-base variants in diploid DNA (C/T, G/A, C/A, G/T, C/G, and T/A) can be grouped into four classes based on the homoduplexes and heteroduplexes formed after PCR (Table 6.1) (Liew et al., 2004). In all cases, heterozygotes are easy to identify by curve shape. Homozygotes in either Class 1 or Class 2 are easy to distinguish by Tm because one homozygote contains an A:T and the other a G:C pair. In short amplicons, the differences in Tm are between 0.8 and 1.4°C for these two classes that make up more than 84% of human single-base variants. Homozygotes in either Class 3 or Class 4 are more difficult to distinguish because the bp (G:C or A:T) stays the same with the bases merely switching strands. Differences in Tm do occur, however, because of nearest-neighbor interactions with adjacent bases, but the differences are generally <0.4°C. Class 3 and 4 variants make up approximately 16% of human single-base variants. Within Classes 3 and 4, nearest-neighbor symmetry occurs in one case out of four (Liew et al., 2004; Palais et al., 2005), predicting identical homozygous melting temperatures in 4% of human single-base variants. Homozygous small insertions/deletions may also be difficult or impossible to

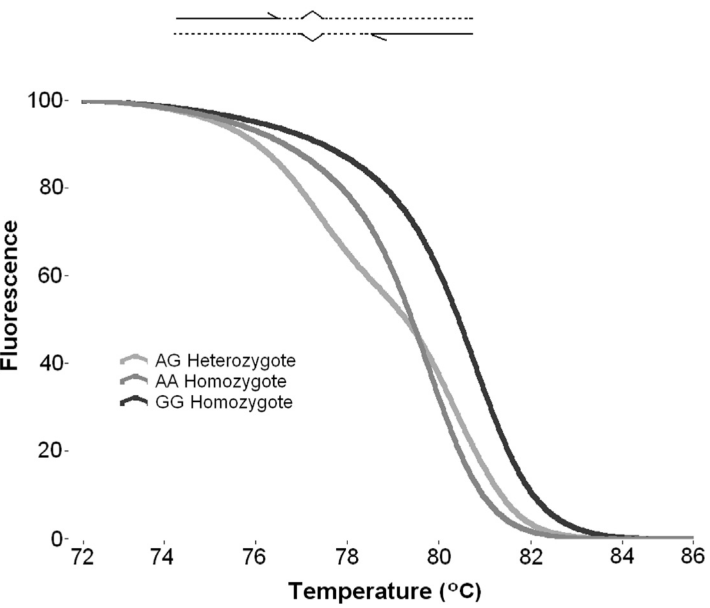


FIGURE 6.4 Genotyping by small amplicon melting. Typical normalized melting curves representing all genotypes of a Class 1, A > G variant are shown. The homozygotes are differentiated by absolute temperature and the heterozygote is discerned by the change in curve shape. Note that no curve overlay is performed for small amplicon genotyping.

TABLE 6.1 Classification of Single-Base Change Variants According to Duplexes (Homoduplexes and Heteroduplexes) Produced After Amplification of a Heterozygote (Liew et al., 2004)

Class	Base Change	Frequency (Human)	Homoduplexes	Heteroduplexes
1	C↔T or G↔A	0.662	C:G and A:T	C::A and T::G
2	C↔A or G↔T	0.176	C:G and A:T	C::T and A::G
3	C↔G	0.088	C:G	C::C and G::G
4	A↔T	0.074	A:T	A::A and T::T

distinguish by small amplicon melting alone: for example, p.F508del in cystic fibrosis (Montgomery et al., 2007). Similarly, in haploid organisms (most microbes) and for hemizygous genes, heterozygotes do not occur and different sequences (strains and genotypes) may have similar or identical Tms. As Tm differences become smaller, instrument precision and solution consistency become more critical for sequence resolution.

There are several ways to discriminate genotypes better with similar Tms. Different alleles can be linked by allele-specific amplification to primer tails that modify amplicon Tm. The tails can be GC- or AT-rich or may incorporate locked nucleic acids (Germer and Higuchi, 1999; Seipp et al., 2008). Internal temperature standards, which increase absolute temperature precision, are particularly useful on plate-based systems (Seipp et al., 2007). For example, most nearest-neighbor symmetric single-base variants (predicted to have identical Tms) can be distinguished by including internal temperature standards that control for both instrument and solution variance (Gundry et al., 2008). However, few instruments incorporate the necessary analysis software required to correct sample temperature changes with internal temperature standards. Finally, unknown samples may be mixed with a known genotype. Typically, the known genotype is wild type. If the unknown is wild type, the melting curve of the mixture will be the same. If the unknown is a homozygous variant, the mixture will be heterozygous and heteroduplexes will alter the curve shape. Mixing can be performed either before or after PCR. If mixing is done after PCR, only the homozygous samples need to be mixed, but the benefits of a closed-tube system are lost. If mixing is done before PCR, quantitative heteroduplex analysis can be performed with optimal mixing ratios to distinguish all three genotypes easily (Palais et al., 2005).

6.2.2 Unlabeled Probe Genotyping

Many sequence variants can be distinguished by amplicon melting alone. However, when greater detail is needed or definitive genotyping is required, unlabeled probes can be used (Zhou et al., 2004a). The same saturating dyes used for amplicon melting can be used for unlabeled probe genotyping. Although high-resolution melting is not required for unlabeled probe genotyping, more genotypes can be distinguished on high-resolution instruments. Asymmetric PCR is performed for 40–50 cycles to produce excess strand complementary to the probe. A 1:5 to 1:10 primer ratio usually produces enough double-stranded product for amplicon melting as well as enough single-stranded product for probe melting (Dujols et al., 2006). The 3′-end of the unlabeled probe is blocked to prevent extension during amplification. This is typically accomplished by 3′-phosphorylation. However, incomplete 3′-phosphorylation may result in probe extension and aberrant melting profiles (Dames et al., 2007). Probe blocking can be improved with amino-modified C6, inverted dT, or a C3 spacer.

Unlabeled probes between 20 and 35 bps with Tms of 55–70°C are generally recommended (Zhou et al., 2004a). If the polymerase lacks 5′ to 3′ exonuclease activity, the probe should melt from all alleles before the PCR extension temperature to prevent possible allele bias and/or PCR inhibition. Higher probe Tms can be used with exonuclease positive polymerases, although it is convenient to keep their Tms below those of primer dimers or other alternative amplification products. Estimated probe Tms are typically 1–4°C lower than observed Tms owing to dye stabilization of the DNA hybrid (Zhou et al., 2005).

After exponential background subtraction (Erali et al., 2008b), unlabeled probe melting data are usually shown on derivative plots to visualize the melting transitions easily. Typical single-base genotyping with an unlabeled probe is shown in Fig. 6.5. Two melting regions are apparent. At lower temperatures, probe melting occurs for specific genotyping. At higher temperatures, amplicon melting occurs. With high-resolution analysis, amplicon melting can be used to scan for variants anywhere between the primers. Genotyping and scanning can even be performed simultaneously from the same melting curve (Zhou et al., 2005).

Probes can be designed to match either the wild-type or variant sequence. For example, common variants revealed by scanning can be definitively identified by matching the probe to the variant (Vandersteen et al., 2007). Probes can

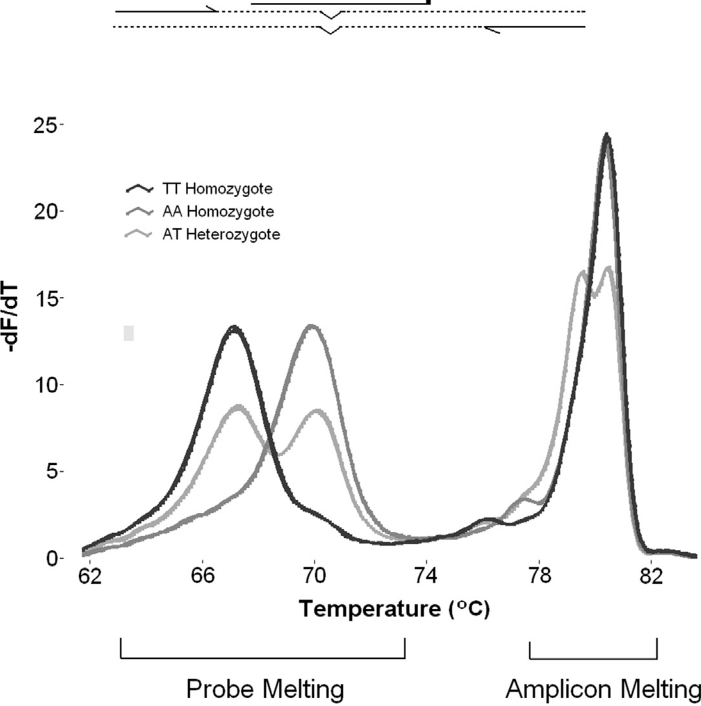


FIGURE 6.5 Genotyping with unlabeled probes. A 3′-blocked unlabeled probe covers the variable region, in this case an A > T variant. After asymmetric PCR, two melting regions are apparent. At low temperature, the probe melts from a single-stranded amplicon, definitively determining genotype. At high temperature the double-stranded amplicon melts, easily identifying the heterozygote but not separating the homozygotes. The small peaks around 76–78°C are extension products of the probes, sometimes seen with incomplete blockage of the probe's 3′-end. The data are displayed as a derivative plot after exponential background removal and normalization.

also be designed to mask benign sequence variations near targeted disease-causing variants by incorporating deletions, unmatched nucleotides, or universal bases (Margraf et al., 2006a). Multiple alleles can be genotyped using unlabeled probes. For example, five variants within exon 10 of the *CFTR* gene were genotyped with two unlabeled probes in the same reaction (Zhou et al., 2005). One of the unlabeled probes genotyped two single-base variants and two deletions.

6.2.3 Snapback Primer Genotyping

Snapback primers are unlabeled probes attached to the 5′-end of a primer. After PCR, the result is a self-probing amplicon in which one PCR strand forms a hairpin (Zhou et al., 2008). Typically, tails are added to the primer that is closer to the variant of interest, so that a smaller hairpin loop is formed. This minimizes additional secondary structures that can confound genotyping. Tail length can be adjusted until a desired hairpin Tm is achieved, which is easily determined by melting. Unlike Scorpion® primers, no covalent modifications are necessary. Asymmetric PCR is usually performed to create intramolecular snapback hairpins as well as intermolecular duplexes of full-length amplicons. Melting of both the amplicon and snapback duplexes allows for investigation of sequence variations within the amplicon as well as variants within the snapback hairpin stem. Similar to unlabeled probe analysis, the melting transitions of both the snapback and full-length amplicon duplexes are usually displayed on derivative plots (Fig. 6.6).

Advantages of snapback primers over unlabeled probes include: (1) only two oligonucleotides are needed instead of three, (2) no 3′-blocking is necessary, and (3) short probe lengths (similar to locked nucleic acids or minor groove binders) can be used because of intramolecular stabilization. This is particularly helpful when an area of interest also contains sequence variants of no clinical significance. The hairpin melting temperature is linearly related to the stem length (6–28 bp) and inversely related to the log of the loop size (17–150 bases) (Zhou et al., 2008). A 2-bp mismatch at the 5′-end of the snapback primer is usually included to limit 3′-extension of the minor snapback product.

6.2.4 Applications of Amplicon Genotyping in Molecular Diagnostics

The simplicity of genotyping by high-resolution melting makes it a favorite for single-locus analysis of sequence

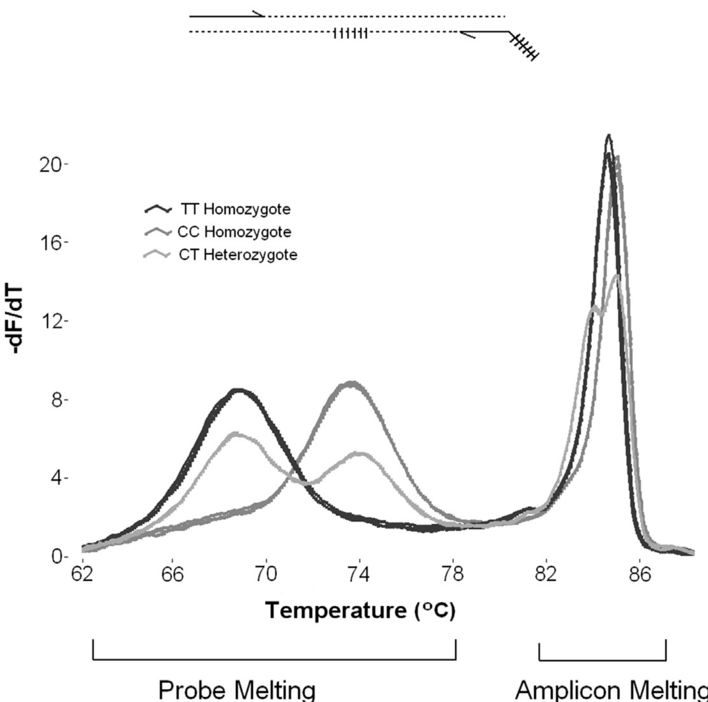


FIGURE 6.6 Genotyping with snapback primers. A snapback primer includes a 5′-tail complementary to its extension product. After asymmetric polymerase chain reaction, regions of single-stranded hairpin (probe) melting and amplicon melting are revealed, similar to unlabeled probe genotyping (Fig. 6.5). High-resolution melting of all genotypes of a C > T variant are shown. The data are displayed as a derivative plot after exponential background removal and normalization.

variation. Genotyping by high-resolution amplicon melting was first reported for common disease variants of *HBB*, *F2*, *F5*, *MTHFR*, *HFE*, and *CFTR* (Gundry et al., 2003; Wittwer et al., 2003; Liew et al., 2004) and now is a commonly used genotyping method used in hundreds of published articles. Amplicon genotyping can be multiplexed if different targets are separated in Tm. Examples include triplex assays for *PAH* (Dobrowolski et al., 2007a) and *OTC* polymorphisms (Dobrowolski et al., 2007b) and a quadraplex reaction for coagulation variants (*F2*, *F5*, and two *MTHFR* loci) that also included two temperature controls for fine calibration (Seipp et al., 2008). Clinical applications of high-resolution amplicon melting in microbiology include mycobacteria speciation using *hsp65* (Odell et al., 2005), bacterial speciation using the 16s rRNA gene (Cheng et al., 2006; Yang et al., 2009), host susceptibility typing for hepatitis C virus (Enache et al., 2015), hepatitis C subtype switching resulting from therapeutic escape (Holysz et al., 2015), *Mycoplasma synoviae* (Jeffery et al., 2007) and varicella zoster (Toi and Dwyer, 2008) strain typing, subtyping influenza A (Lin et al., 2008), and many others. For methicillin-resistant *Staphylococcus aureus*, small amplicon melting may replace sequence typing for epidemiologic tracking (Mongelli et al., 2015; Mayerhofer et al., 2015; Mazi et al., 2015).

Unlabeled probes have been used to genotype *F5* and *CFTR* (Zhou et al., 2004a, 2005), *LCT* (Janukonyte et al., 2010), and multiple *MBL2* genotypes (Vossen et al., 2010). Mutations were identified in the *RET* protooncogene with unlabeled probes that masked common polymorphisms (Margraf et al., 2007). Unlabeled probes have also been used to detect and differentiate herpes simplex virus (HSV)-1 and HSV-2 (Cho et al., 2008), drug resistance mutations in cytomegalovirus (Chen et al., 2015), human papillomavirus genotypes (Lee et al., 2012), and many other variants in human, plant, bacterial, and viral organisms.

Snapback primers have been used to genotype *F5* c.1601 G > A, TA repeats in Gilbert syndrome (Farrar et al., 2011), and several *CFTR* variants (Zhou et al., 2013). Snapback primers have also been used for the early detection and genotyping of zoonotic viruses of the *Capripoxvirus* genus (Gelaye et al., 2013).

6.3 VARIANT SCANNING BY HIGH-RESOLUTION MELTING

Variant scanning by high-resolution melting depends on detecting heteroduplexes after amplification. After PCR of a heterozygous sequence, four unique duplexes are formed: two homoduplexes (in which both strands are completely

complementary) and two heteroduplexes (with at least one mismatch between strands). The two homoduplexes are identical to the starting DNA and have Tms that are usually close to each other. The remaining two duplexes are heteroduplexes with mismatched bases and lower Tms. The observed melting curve is a composite of the melting profiles of these four duplexes. Other scanning techniques also depend on detection of heteroduplexes, but high-resolution melting is unique because no physical separation of heteroduplexes is necessary.

Variant scanning by high-resolution melting uses a saturating DNA dye to detect heteroduplexes. The dye is usually included in the PCR buffer so that product accumulation can be optionally monitored each cycle in real time. After PCR, the composite melting curve of two homoduplexes and two heteroduplexes is measured by high-resolution melting. Heteroduplexes are detected by a change in melting curve shape after normalization (Fig. 6.3B) and curve overlay (Fig. 6.3C). Although these differences are present on derivative plots (Fig. 6.3D), they are easier to identify on difference plots (Fig. 6.3E). Heteroduplex detection by high-resolution melting is favored by rapid cooling before melting, rapid heating during melting, and low Mg^{2+} concentrations (Gundry et al., 2003). Depending on the instrument and protocol used, high-resolution melting can be performed in under 1 min or may take longer than 90 min (Herrmann et al., 2007b; Sundberg et al., 2014).

The sensitivity and specificity of high-resolution melting to detect heterozygous single-base changes was comprehensively evaluated using engineered plasmids (Reed and Wittwer, 2004). The effects of PCR product size, type of base change, and GC content, and the position of the base change in relation to the PCR product were considered. The PCR products ranged in size from 50 to 1000 bp with GC contents of 40–60%. Sensitivity and specificity for products <400 bp were 100% ($n = 576$). For products between 400 and 1000 bp the sensitivity was 96.1% and the specificity was 99.4% ($n = 672$). The location or identity of the single-base change did not affect sensitivity.

Statistics for published studies that targeted variant scanning of inherited disease were last summarized in 2010 (Montgomery et al., 2010). Up to that time, 29 studies reported data adequate to calculate high-resolution melting scanning sensitivity. Most (23 of 29) reported 100% heterozygote sensitivity. The weighted average sensitivity was 99.3% ($n = 839$) and the specificity was 98.8% ($n = 2659$). The PCR product size varied from 79 to 658 bp.

High-resolution melting is unusual among scanning techniques in that most homozygous variants can be identified. The weighted average sensitivity for detecting homozygous variants in nine studies was 91.6%, although 79% of these cases came from one study (Dobrowolski et al., 2009). To detect all homozygous and hemizygous variants, mixing with a known genotype is necessary either before or after PCR (Liew et al., 2004; Palais et al., 2005). DNA is usually extracted from whole blood for genetic scanning studies, although dried blood spots (Dobrowolski et al., 2007a) whole-genome amplification (Cho et al., 2008), and whole-genome amplification of dried blood spots (Margraf et al., 2006b; Winkel et al., 2011) all have been reported.

The sensitivity and specificity of scanning are usually decreased when tumor tissue is analyzed because of normal cell admixture and/or sample fixation/staining before DNA extraction. Summary statistics before 2010 (Montgomery et al., 2010) analyzed 26 studies of high-resolution melting used to scan for variants in tumors, typically from formalin-fixed, paraffin-embedded tissue. Most (21 of 26) reported 100% heterozygote sensitivity. The overall heterozygote sensitivity was 96.4%, comprising 642 positive results. Amplicon size varied from 51 to 653 bp. The minimum allele fraction was 0.1–20% and varied with the PCR product size and the target.

Common heterozygous variants of no clinical significance (polymorphisms) are a problem for any scanning method. Sometimes benign intronic variants can be avoided by placing the primers closer to intron–exon boundaries, being careful to bracket frequent splice site positions (Erali and Wittwer, 2010). If common variants cannot be avoided by primer placement, deoxyinosine can be incorporated into the primer at benign variant positions to minimize allele-specific amplification (Hondow et al., 2011). If the benign variants are within the exon or nearby intronic splice regions, common variant clusters identified by screening a normal population are seldom of concern because the melting curves of about 90% of heterozygous variants are different from each other (Wittwer, 2009). However, to be certain the variant is common and not a rare variant that could be pathogenic, genotyping by small amplicon melting (Dobrowolski et al., 2007a,b), unlabeled probes (Vandersteen et al., 2007), or snapback primers (Zhou et al., 2008) can almost always avoid the need for sequencing. For example, snapback primers can be used elegantly to both scan and genotype (Zhou et al., 2013). After symmetric PCR and melting to obtain a variant scan of the entire PCR product, the reaction is diluted with water to promote intracellular snapback formation. A second melting curve then confirms the genotype of common variants, leaving only very rare variants for sequencing. These methods have been applied to *CFTR*, reducing the need for sequencing PCR products to <1 in 200.

6.3.1 Applications of Variant Scanning in Molecular Diagnostics

Among nonsequencing scanning methods, high-resolution melting is clearly favored over legacy methods that require separations (denaturing high-performance liquid

chromatography, single-strand conformation polymorphism, denaturing gradient gel electrophoresis, etc.) across most metrics, including cost, time, simplicity, and sensitivity/specificity. Many variants missed by prior methods can be revealed by high-resolution melting (Out et al., 2015). Comparison against Sanger sequencing is less clear for single genes, but arguments have been made in favor of melting (Zhou et al., 2013). Massively parallel sequencing has obvious advantages over melting if many genes need analysis, because the sheer volume of individual targeted reactions becomes unmanageable. Hundreds of publications have used high-resolution melting to scan genes for variants. Examples include BRCA1/2 (Hondow et al., 2011) and nondeletion variants in alpha-thalassemia (*HBA1/2*) (Petropoulou et al., 2015) and Duchenne–Becker muscular dystrophy (*DMD*) (Esterhuizen et al., 2014). Many disease genes such as *BRCA1/2*, *CFTR*, *F8*, *c-kit*, *EGFR*, and *TP53* have been studied by multiple groups (Montgomery et al., 2010).

In addition to heterozygote scanning of diploid organisms, haploid microbes can be studied by mixing with a known sample. For example, mixing has been used for bacterial speciation using the 16S ribosomal gene (Cheng et al., 2006; Yang et al., 2009) and antibiotic resistance of mycobacteria using *rpoB* (Hoek et al., 2008), and influenza A can be subtyped using the *M* gene (Lin et al., 2008). Another interesting use of mixing is to locate RNA editing sites by mixing cDNA with DNA, as demonstrated in *Arabidopsis* (Chateigner-Boutin and Small, 2007).

6.4 SPECIFIC EXAMPLES OF HIGH-RESOLUTION MELTING IN CLINICAL DIAGNOSTICS

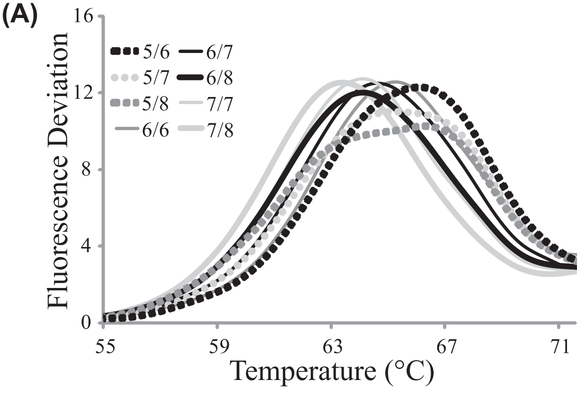
6.4.1 Factor V Leiden Genotyping

Factor V Leiden is a Class 1 single-nucleotide variant (c.1601 G > A) that increases the risk of deep venous thrombosis and pulmonary embolism. Genotyping by melting was first performed with labeled hybridization probes. However, high-resolution melting eliminates the need for labeled probes and enables genotyping with small amplicons, unlabeled probes, or snapback primers. Small amplicon melting is the simplest method and can be performed on most instruments because the homozygous genotypes are well separated by temperature (about 1°C) and the heterozygous shape difference is large when saturation dyes are used (Liew et al., 2004). Rare, unexpected nearby heterozygotes (one single-base change, one single-base deletion, and one compound heterozygote) could also be distinguished from the wild-type sequence and Factor V Leiden (Graham et al., 2005). Simultaneous amplicon and unlabeled probe analysis of Factor V Leiden was later shown in 384-well format (Zhou et al., 2005). Both probe and amplicon melting provided independent assessments of genotype for increased confidence. Factor V Leiden genotyping has also been demonstrated with snapback primers (Zhou et al., 2008).

Small amplicon genotyping of Factor V Leiden, as well as other thrombophilia-associated variants, has been achieved by rapid serial PCR and high-speed melting (0.5°C/s) on a microfluidic platform (Sundberg et al., 2014). The instrument also performs automated liquid handling for PCR setup and execution. Typical run times for PCR and melting of 50-nL reactions are 12.5 min. Two study sites each performed a blinded study on 100 clinical samples enriched for *F2* c.*97G > A, *F5* c.1601 G > A and *MTHFR* c.665 C > T and c.1286 A > C variants. Multiple, serial testing and automated liquid handling allowed sequential PCR and melting for each clinical sample. Internal temperature controls (Seipp et al., 2007) were used to improve Tm precision and genotyping accuracy. Results at both sites were concordant with an instrument accuracy of 100%. One unique feature of the testing platform is the ability to perform automatic reflex or repeat testing depending on prior test results. The automated liquid handling also reduces time and the potential for human error during manual PCR setup. The instrument has also been used for *CFTR* mutation scanning and targeted genotyping with unlabeled probes (Cao et al., 2015).

6.4.2 Gilbert Syndrome Genotyping

Gilbert syndrome is a chronic, nonhemolytic, unconjugated hyperbilirubinemia associated with increased thymine-adenine (TA) repeats within the promoter of *UGT1A1*. Although usually of limited clinical significance (Bosma et al., 1995), Gilbert syndrome is associated with an increased risk for drug toxicity during cancer chemotherapy with irinotecan (Innocenti et al., 2004). This association resulted in safety relabeling and a recommendation for genotype-guided dosing (Hoskins et al., 2007; Minucci et al., 2010). Small amplicon genotyping of the $(TA)_n$ promoter polymorphism of *UGT1A1* can resolve all genotypes from $(TA)_5$, $(TA)_6$, and $(TA)_7$ alleles (Thomas et al., 2013). However, when $(TA)_8$ alleles were included, complete genotype resolution required a snapback primer (Farrar et al., 2011). $(TA)_5$ and $(TA)_8$ alleles are common in black African populations. Snapback primer genotyping of the *UGT1A1* $(TA)_n$ promoter polymorphism is shown in Fig. 6.7. A fluorescence background removal method was used to plot the local deviation from exponential decay, which improved hairpin melting display and genotype clustering (Fig. 6.8). In a blinded study of 100 African American DNA samples, snapback primer genotyping using a capillary-based melting instrument correctly identified all genotypes. Instrument temperature precision was critical for accurate genotyping, with accuracy falling to 81% when



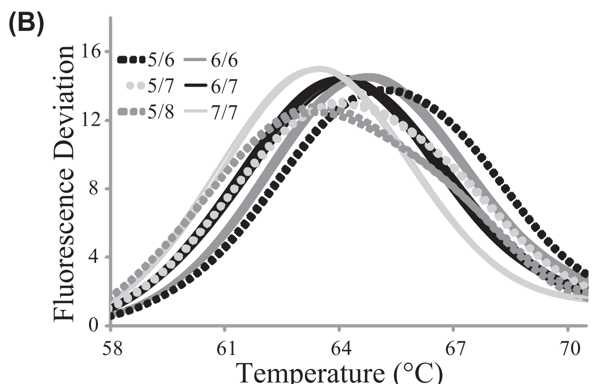


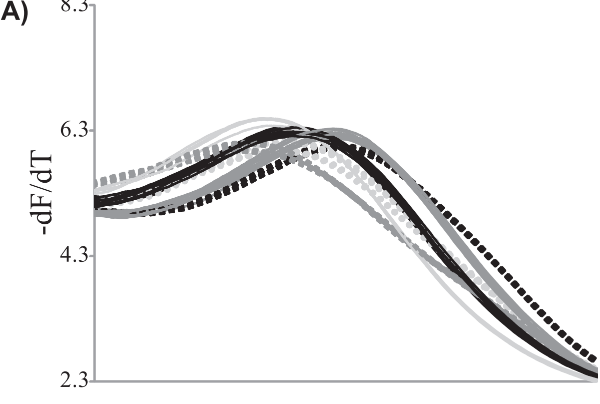
FIGURE 6.7 Snapback primer genotyping of the *UGT1A1* $(TA)_n$ promoter polymorphism. (A) Subtle differences in 8 common genotypes, although small, allow accurate genotyping of the $(TA)_n$ promoter polymorphism. (B) Snapback primer deviation melting curves for multiple samples in a typical run. Numbers indicate the $(TA)_n$ repeat genotype: 5/6 (2 samples), 5/7 (4 samples), 5/8 (2 samples), 6/6 (8 samples), 6/7 (10 samples), and 7/7 (4 samples).

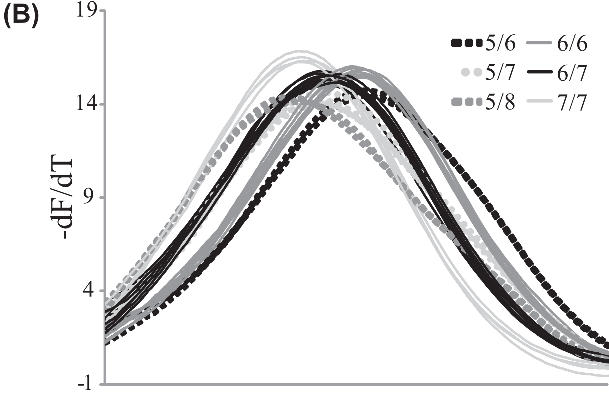
the assay was performed on a plate-based melting instrument. The limits of small amplicon genotyping for the $(TA)_n$ promoter polymorphism were also confirmed, with an overall error rate of 16%. When greater sequence resolution is needed, snapback primer genotyping can improve genotyping of dinucleotide repeats. High-resolution melting was also used to scan for single-nucleotide changes in the promoter and 11 exons of *UGT1A1* in a Taiwanese population (Tsai et al., 2014).

6.4.3 Cystic Fibrosis Variant Scanning

Cystic fibrosis is an autosomal recessive disorder caused by mutations in *CFTR* affecting the exocrine glands of the lungs, pancreas, intestines, and liver (Riordan et al., 1989). Over 2000 variants of the *CFTR* gene have been described, each with varying frequencies and distribution across populations.

Complete analysis of all 27 exons of *CFTR* was first demonstrated by high-resolution melting in 2007 (Montgomery et al., 2007). Initially, common variants were identified in 96 Caucasian blood donors and correlated with their melting patterns. Then, 30 blinded samples enriched for disease variants were used to assess the sensitivity of the technique. Scanning detected all 40 disease-associated heterozygotes for a sensitivity of 100%. However, 47 benign heterozygotes were also detected. By considering the melting patterns of the six most common heterozygotes,





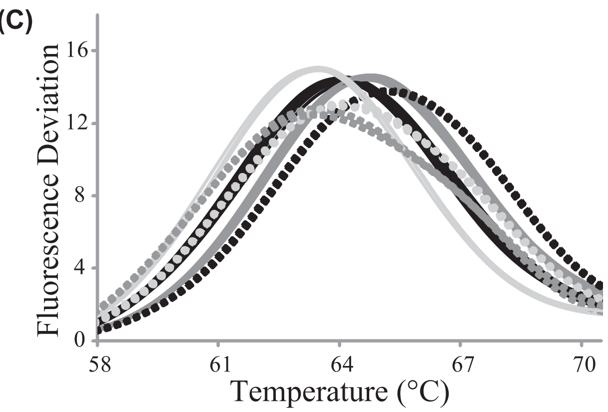


FIGURE 6.8 Comparison between different analysis methods for snapback primer genotyping of the *UGT1A1* $(TA)_n$ promoter polymorphism. (A) Derivative melting curves before normalization and exponential background removal. (B) Derivative melting curves after normalization and exponential background removal. (C) Deviation melting curves. Genotype clustering is improved and hairpin melting is better visualized with deviation analysis.

45 of these 47 were matched to benign variants and did not require sequencing or genotyping. Most (6 of 8) homozygous variants were also distinguished from wild-type without mixing or genotyping.

In an extension and expansion of this work, snapback primers were used for sequential scanning and genotyping of variants within the *CFTR* gene (Zhou et al., 2013). All 27 exons and neighboring splice sites were amplified and specific snapback primer genotyping was performed on all 23 variants recommended by the American College of Medical Genetics, as well as seven common variants. Symmetric PCR was used and the products were initially melted to scan for variants, and subsequently diluted with water to encourage intramolecular hairpin formation for genotyping by melting, as outlined in Fig. 6.9.

When known mutations occurred close together, one snapback primer was used to genotype both mutations, as demonstrated for p.G551D and p.R553X. Variants that were contained in the same amplicon but were separated by a large intervening sequence could be genotyped using a dual snapback approach with both forward and reverse primers having snapback tails. This was used on exon 11 in which the common c.1408 A > G variant was genotyped with one snapback primer and p.F508del and p.I507del with another snapback primer. A single snapback primer can also be used to cover separated variants by introducing a bulge in the PCR product under the snapback tail. Variants p.W1282X and c.2870 A > G, which are separated by 23 bp, were genotyped with one snapback primer by inducing a 10-bp bulge in the PCR product. This was accomplished by removing a 10-bp sequence between the two variants in the complementary snapback tail. Improved Tm predictions for bulge-inducing snapback tails, as well as software tools to automate the snapback design, should promote greater use of this genotyping approach.

6.4.4 Mitochondrial Genome Variant Scanning

Mitochondria are unique organelles that have a circular genome of approximately 16.5 kb that encodes genes for independent protein synthesis, as well as the cellular machinery for generating adenosine triphosphate through oxidative phosphorylation (Dimauro and Davidzon, 2005). Unlike nuclear DNA, which has two copies per cell, mitochondrial DNA can exist as hundreds to thousands of copies per cell. Comprehensive mitochondrial genome profiling and targeted genotyping of known pathogenic mutations has been performed by high-resolution melting (Dobrowolski et al., 2009). This impressive task used 36 overlapping amplicons and could be completed in 1 h. Heteroplasmic variants were detected at or below 10% allele fraction and 96.5% of homoplasmic variants (no heteroduplex formed) were detected during scanning. Targeted small amplicon genotyping detected causative mutations for myopathy, encephalopathy, lactic acidosis, and stroke-like episodes (m.3243 A > G), myoclonus, epilepsy, and ragged red fibers (m.8344 A > G), and aminoglycoside antibiotic-induced hearing loss (m.1555 A > G). Comprehensive mutation scanning of mitochondrial DNA further demonstrates the utility of high-resolution melting for novel variant discovery and genotyping.

6.5 OTHER APPLICATIONS OF HIGH-RESOLUTION MELTING IN MOLECULAR DIAGNOSTICS

In addition to genotyping of known variants and scanning for unknown variants, methylation analysis, rare allele detection, copy number analysis, and sequence matching are important applications of high-resolution melting.

6.5.1 Methylation Analysis

The methylation of cytosine in DNA is important in the regulation of gene activity and expression (Breiling and Lyko, 2015). Methylation of cytosine (C) to 5-methylcytosine (m^5C) typically occurs within 5′−C−phosphate−G−3′ islands near the promoters of protein-encoding genes. Methylation information is lost during PCR, so either methylation-specific restriction endonucleases or bisulfite treatment is used before amplification. Bisulfite treatment converts Cs to uracil, whereas m^5Cs are not affected, allowing methylation-specific PCR that is sensitive to methylation under primer binding sites and/or methylation-specific high-resolution melting that reveals the methylation dosage of PCR products between the primers (Wojdacz and Dobrovic, 2007).

Bisulfite treatment and PCR convert unmethylated C:G pairs to A:T pairs of lower stability so that the PCR product Tm is directly related to the degree of methylation. In contrast to methods that perform best with heteroduplex detection, saturation dyes are not required or desired; methylation analysis by melting was introduced with SYBR® Green I (Worm et al., 2001). High-resolution melting appears to increase the sensitivity and precision of analysis (Wojdacz and Dobrovic, 2007; Dahl and Guldberg, 2007). Variations include melting analysis after real-time methylation-specific PCR (Kristensen et al., 2008) and digital methylation-sensitive high-resolution melting (Snell et al., 2008).

Methylation analysis by high-resolution melting can diagnose imprinting disorders, including Angelman and Prader−Willi syndromes (White et al., 2007) and Beckwith Wiedemann and Russell Silver syndromes (Alders et al., 2009). Studies on promoter methylation include *BRCA1*

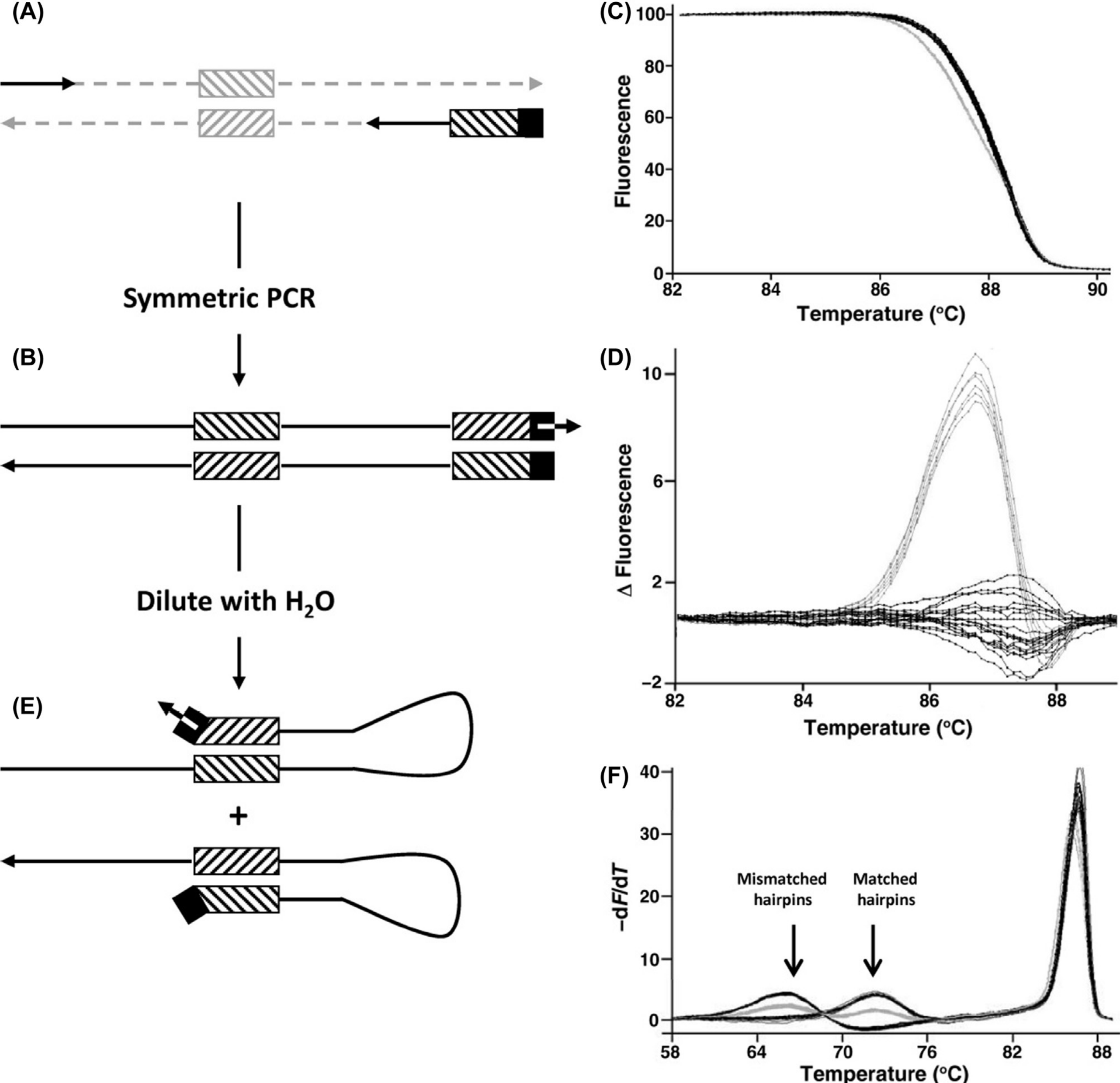


FIGURE 6.9 Snapback primer scanning and genotyping using symmetric polymerase chain reaction (PCR) and product dilution. (A) 5′-tailing (diagonal cross-hatch) of a PCR primer produces a snapback primer that is complementary to its extension product and probes a region of interest. Snapback primers include an added 2-base pair sequence (*solid black*) that is not complementary to the area of interest and prevents hairpin extension. (B) Symmetric amplification generates a PCR product that includes the complementary tail sequence. Duplex, intermolecular PCR hybridization is favored at the end of PCR owing to high complementary strand concentrations and solution ionic strength. (C) Small amplicon melting of *CFTR* exon 27 after symmetric PCR using a snapback primer targeted to the common c.4389 G > A variant. After exponential background subtraction, normalization, and curve overlay, most products cluster together (*black*), whereas some products are easily identified by a difference in curve shape (*gray*). (D) The same data displayed as a difference plot. (E) After a fivefold dilution with water and product denaturation, the reduced ionic strength and lower strand concentrations favor partial formation of intramolecular hairpins. (F) Repeat melting analysis reveals both intermolecular (amplicon) and intramolecular (hairpin) structures. Hairpin stability is greater for matched hairpins than for mismatched hairpins, allowing complete genotyping of the sequence around the common c.4389 G > A variant.

(Huang et al., 2015), *HIN-1*, *MGMT*, *RASSF1A* (Spitzwieser et al., 2015), and many others. *MGMT* and *BNIP3* methylation down to 0.1−1% could be detected in colorectal cancer samples (Wojdacz and Dobrovic, 2007).

6.5.2 Rare Allele Detection

Snapback primers can be used for minor allele enrichment and detection (Zhou et al., 2010). Direct sequencing is generally limited to the detection of allele fractions of 10−20% (van den Oever et al., 2015). However, the relative proportion of mutant to wild-type alleles in tumor samples is often below this detection limit. Using snapback primers to block amplification of wild-type alleles, detection limits of 0.1% for *BRAF* p.V600 E and 0.02% for *EGFR* exon 19 in-frame deletions were achieved. An advantage of this method is that not only are rare alleles enriched, they are also genotyped by snapback primer probe Tms. Minor allele enrichment was found to increase with shorter extension times, reduced Mg^{2+} concentrations, and higher probe Tms. Snapback primer allele enrichment has been independently extended to *KRAS* and *EGFR* (Sun et al., 2014) and *JAK2* and *MPL* mutations (Wu et al., 2014a). High-resolution melting can also be used with allele-specific PCR and/or blocking probes for sensitive detection of rare variants (Smith et al., 2011) or paternally inherited mutations in maternal plasma (van den Oever et al., 2015).

When the rare allele sequence is not known or when changes are spread out over long distances, high-resolution melting can be combined with other techniques. For example, rare allele detection can be achieved by combining digital PCR with high-resolution melting. Using a combined digital PCR and high-resolution melting scanning approach, somatic mutations associated with colorectal cancer (*KRAS*, *BRAF*, *APC*, and *TP53*) were detected down to 0.1% in DNA obtained from stool samples (Zou et al., 2009). Positive reactions with melting curves different from wild-type were further characterized by sequencing. This approach was also used to increase the sensitivity of sequencing for detection of rare mutations found in non−small cell lung cancer (Do and Dobrovic, 2009). Co-amplification at lower denaturation-temperature PCR has been combined with high-resolution melting of small amplicons (Pang et al., 2013) or unlabeled probes (Wu et al., 2014b) to improve sensitive detection of rare variants. Competitive amplification of differentially melting amplicons followed by high-resolution melting is another combination to detect rare variants (0.025−0.25%) of all types within a PCR product (Kristensen et al., 2012).

6.5.3 Copy Number Analysis

Large-scale genomic projects such as HapMap and the 1000 Genomes Project have demonstrated that structural variation in the human genome is widespread. Copy number variation is increasingly associated with complex genetic diseases such as cancer, heart disease, diabetes, and mental health disorders (Ono et al., 2015; Cui et al., 2015). These associations have been fueled by large data sets generated by impressive discovery tools such as microarrays (Carter, 2007) and massively parallel sequencing (Zhao et al., 2013). However, once associations are confirmed, targeted analysis of copy number variation is desirable by methods that are simple and inexpensive. High-resolution melting has been used to determine copy number variation at the homologous genes *SMN1* and *SMN2*, where the copy number at each locus influences the phenotype of spinal muscular atrophy (Dobrowolski et al., 2012; Er et al., 2012; Morikawa et al., 2011). Competitive PCR, in which similar sequences are amplified on different chromosomes with one primer pair, can be combined with high-resolution melting for relative quantification of common aneuploidies (Guo et al., 2012). When sequence variants or common primers cannot be used to determine relative copy number, PCR can be terminated during the exponential phase and copy number determined by melting relative to a reference product on a separate chromosome. Copy number changes associated with the *APC* gene in familial adenomatous polyposis and the X-linked *DMD* gene in Duchenne muscular dystrophy have been determined this way (Borun et al., 2014).

A simpler alternative to limiting PCR by cycle number is to limit the amount of deoxynucleotide triphosphates (dNTPs) in the PCR (Zhou et al., 2015). By limiting dNTPs, PCR no longer needs to be stopped during the exponential phase because the dNTPs are consumed before any primer-specific changes in PCR efficiency can occur. That is, the efficiencies of both PCR products (target and reference) are both affected equally as the dNTPs are consumed. Using limiting dNTPs, copy number changes of common trisomies (13, 18, and 21), large exonic deletions (*CFTR*), gene amplifications (*EGFR*), sex chromosome abnormalities, and spinal muscular atrophy (*SMN1/2*) were demonstrated. For best accuracy, the reference melting peaks are first normalized for both fluorescence and temperature differences, after which relative copy number changes can be visualized directly by comparison with controls (Fig. 6.10). Initial studies suggest that the technical variation of copy number analysis by limiting dNTPs (coefficient of variation of 1%) is better than that typically obtained with a 20,000 event digital PCR (Zhou et al., 2015; Dobrovic, 2015). Limiting dNTPs for relative copy number assessment is a simple and precise way to confirm copy number variation.

6.5.4 Sequence Matching

In some cases, specific genotyping is not as important as knowing whether the same sequence is present. Interesting

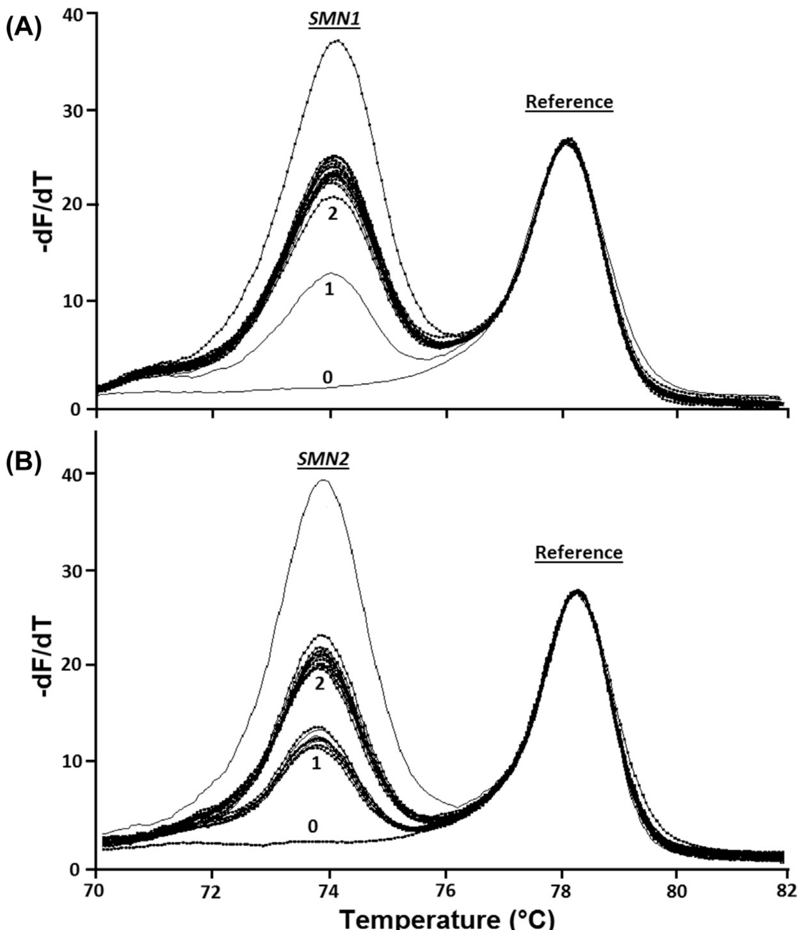


FIGURE 6.10 Quantification of: (A) *SMN1* and (B) *SMN2* by competitive amplification against a reference gene using restricted deoxynucleotide triphosphates. The proportion of genotypes from 24 normal samples approximates the expected distribution: 2 *SMN1*/2 *SMN2* (58% observed vs. 57% expected), 2/1 (33% observed vs. 28% expected), 2/0 (4% observed vs. 3% expected), and 3/2 (4% observed vs. 3% expected). Single known samples were used to demonstrate the *SMN1* = 0 and *SMN1* = 1 copy genotypes. *SMN1* and *SMN2* are on chromosome 5 whereas the reference gene is on chromosome 7.

examples are HLA analysis for transplantation compatibility, repeat typing for identity matching, and genetic mapping.

6.5.4.1 Human Leukocyte Antigen Matching

Conventional HLA testing requires specific typing of six highly polymorphic loci, an onerous task. In contrast, high-resolution melting can easily establish HLA sequence identity and transplant compatibility, at least for living related transplants. For example, high-resolution melting was applied to the highly polymorphic HLA-A locus (Zhou et al., 2004b). All seven cases of shared alleles between two individuals were assessed and a 17-member family was tested. Results from melting curve analysis were concordant with sequencing. HLA identity was suggested when two individuals had the same melting curve. Identity was confirmed by comparing the melting curve of a 1:1 mixture with the original individual curves. This technique can be extended to other HLA loci for more complete HLA matching. Analysis of unrelated donor–recipient pairs for allogeneic hematopoietic stem-cell transplantation is a more difficult but intriguing possibility.

6.5.4.2 Repeated Sequences

Repeat sequence polymorphisms are usually typed by size separation on gels. Several studies have used high-resolution melting analysis to differentiate these polymorphisms, such as 21-bp tandem repeats in the promoter of *XRCC5* (Rajaei et al., 2012) and short tandem repeats for forensics (Nguyen et al., 2012; Nicklas et al., 2012). Profiling of deletion–insertion polymorphisms after allogeneic stem cell transplantation can detect the persistence or reappearance of host cells (Gerini et al., 2014). Differentiation of clustered short sequence repeats of *Campylobacter jejuni* (Price et al., 2007), the complex repeat structure of *spa* in *S. aureus* (Stephens et al., 2008), and multilocus short repeats in *Bacillus anthracis* (Fortini et al.,

2007) have been reported. High-resolution melting of dinucleotide repeats can also identify varieties of grapes and olives (Mackay et al., 2008).

6.5.4.3 Genetic Mapping and Editing

Single-base variants can be used in genetic mapping when the sequence and position of the variant are known. In addition, high-resolution melting can provide fine mapping in the absence of specific genotyping (Lehmensiek et al., 2008). High-resolution melting is frequently selected as the method of choice in plant gene mapping to detect and genotype single-nucleotide variants rapidly and cost-effectively (Chagne, 2015). Paired with new genome editing methods (transcription activator-like effector nucleases and clustered regularly interspaced short palindromic repeats), high-resolution melting is a simple and efficient way to identify novel mutant alleles (Thomas et al., 2014; Wang et al., 2015b).

6.6 MELTING CURVE PREDICTION AND ASSAY DESIGN TOOLS

In an attempt to democratize high-resolution melting, a number of interactive design and analysis tools are available free of charge and without registration at https://www.dna.utah.edu. These tools include uMelt (Dwight et al., 2011), uMeltHETs (Dwight et al., 2014), uDesign, and uAnalyze (Dwight et al., 2012), among others. We hope to promote an educated and resourceful community of expert users and welcome feedback on these software tools.

uMelt was developed to provide an easy-to-use Web-based tool that can predict DNA melting curves and facilitate the design of high-resolution melting assays (Dwight et al., 2011). This tool is particularly useful for predicting and confirming PCR product melting curves, especially amplicons with multiple melting domains. Traditional wisdom concerning product verification by melting led to the common misconception that "pure" PCR products must

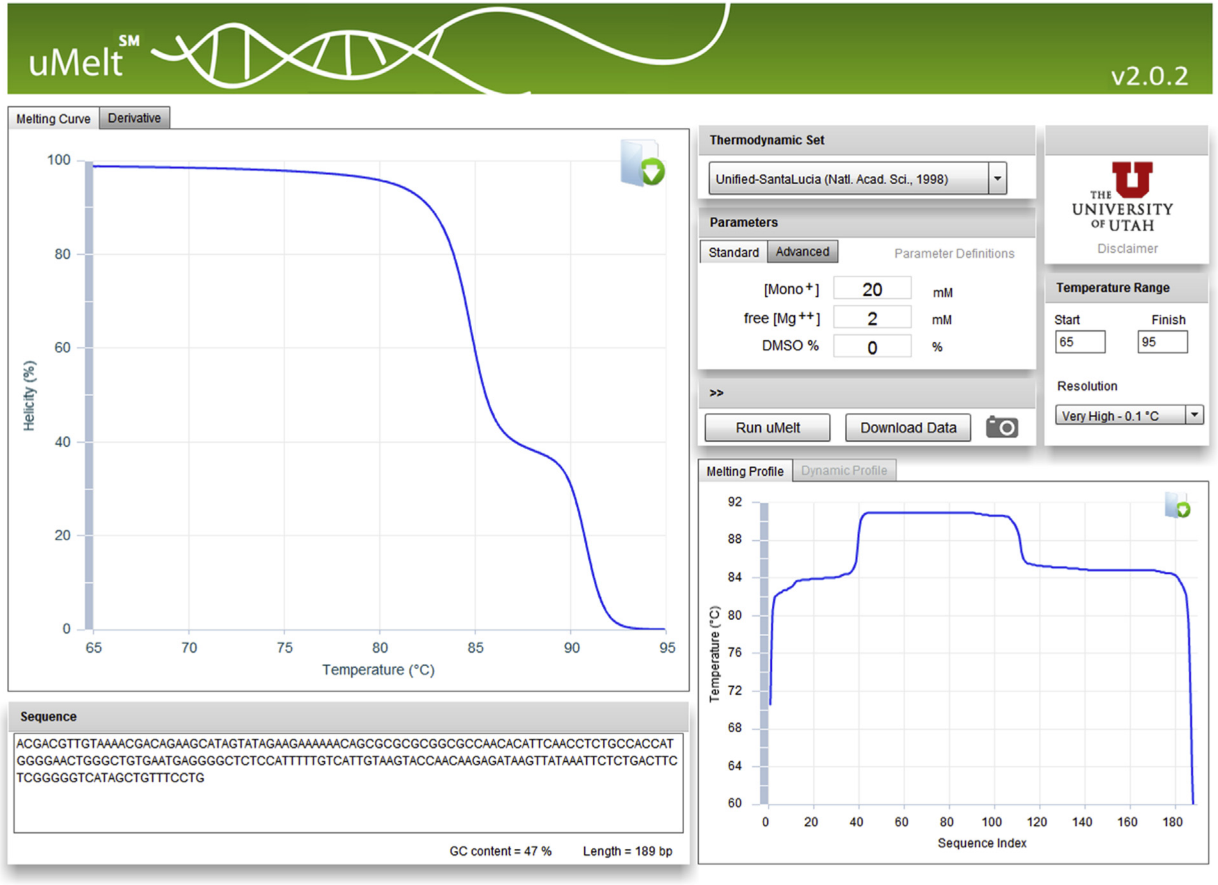


FIGURE 6.11 Polymerase chain reaction product melting curve prediction. Shown is an uMelt user interface displaying the predicted melting curve and melting profile for a defined amplicon sequence. The user copies the desired sequence to bottom left and adjusts parameters at top right as necessary. The sequence shown is 189 base pairs in length and has three melting domains (*lower right*), of which only two are seen on background subtracted, normalized melting plots (*upper left*). *DMSO*, dimethyl sulfoxide.

display only one melting transition. Although for 40- to 100-bp PCR products this is often true, 200- to 600-bp products usually differ in GC content and stability across the amplicon, resulting in multiple melting domains. With uMelt, DNA melting can be visualized as a "dynamic profile" that reveals sequence-specific melting domains as the temperature is increased. During initial assay design it is helpful to predict the temperature at which the product is completely melted, so that the denaturation temperature can be raised if necessary. Primers that generate high-Tm products can be redesigned or Tm depressors (DMSO or betaine) can be added to the PCR. If the PCR product melts between 80°C and 90°C, the denaturation temperature can be reduced to speed amplification and increase specificity. The front panel of uMelt is shown in Fig. 6.11.

Small amplicon genotyping is particularly attractive for most single-nucleotide variants. uDesign prompts users to enter a National Center for Biotechnology Information reference single nucleotide polymorphism number (www.ncbi.nlm.nih.gov/snp) for the variant of interest and a primer pair is automatically generated based on the desired primer Tm. Predicted melting curves for all three genotypes of the amplified product are also displayed. For any particular variant, when small amplicon genotyping is not sufficient to distinguish the homozygotes or when greater separation of genotypes is desired, an unlabeled probe or snapback primer can be designed. uMelt probes predict the probe Tm and resulting melting curves for genotyping. Finally, uAnalyze (Dwight et al., 2012) displays raw data from many instruments, removes the fluorescent background, and normalizes experimental data, with additional options of curve overlay and difference plots to identify heterozygotes. With user input of the product sequence, experimental results can also be compared with the predicted Tm for visual verification of the amplified product.

6.7 CONCLUSIONS

High-resolution melting can be used for genotyping, variant scanning, sequence matching, methylation analysis, rare allele detection, and copy number analysis. The amount of information that can be extracted from simple melting of PCR products entirely depends on the resolution and precision of the melting curves, which in turn depends on the available instrumentation and software. Dyes that detect heteroduplexes increase the sensitivity of scanning and the accuracy of genotyping applications but are not necessary for many other applications, including methylation analysis and repeat typing. Genotyping with unlabeled probes or snapback primers can be performed on standard instrumentation, but the resolution of multiple genotypes is improved with better instrumentation and software.

To quote a 2009 review (Vossen et al., 2009), "… High resolution melting analysis is a multipurpose technology and standard tool that should be present in any laboratory studying nucleic acids." High-resolution melting methods are fast, affordable, flexible, and simple. DNA melting will be around as long as life as we know it exists. As long as instruments and methods continue to improve, additional applications will appear. For example, the FilmArray® is a 1-h sample-to-answer multiplex microbial detection platform that uses high-resolution melting for detection (Poritz et al., 2011). In the future, pairing high-resolution melting with extreme PCR (Farrar and Witter, 2014) and temperature control by fluorescence (Sanford and Wittwer, 2014) should result in decreased time and increased resolution for further melting applications.

REFERENCES

Alders, M., Bliek, J., vd Lip, K., vd Bogaard, R., Mannens, M., 2009. Determination of KCNQ1OT1 and H19 methylation levels in BWS and SRS patients using methylation-sensitive high-resolution melting analysis. Eur. J. Hum. Genet. 17 (4), 467−473.

Bernard, P.S., Ajioka, R.S., Kushner, J.P., Wittwer, C.T., 1998. Homogeneous multiplex genotyping of hemochromatosis mutations with fluorescent hybridization probes. Am. J. Pathol. 153 (4), 1055−1061.

Borun, P., Kubaszewski, L., Banasiewicz, T., et al., 2014. Comparative high resolution melting: a novel method of simultaneous screening for small mutations and copy number variations. Hum. Genet. 133 (5), 535−545.

Bosma, P.J., Chowdhury, J.R., Bakker, C., et al., 1995. The genetic basis of the reduced expression of bilirubin UDP-glucuronosyltransferase 1 in Gilbert's syndrome. N. Engl. J. Med. 333 (18), 1171−1175.

Breiling, A., Lyko, F., 2015. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. Epigenet. Chromatin 8, 24.

Bruzzone, C.M., Steer, C.J., 2015. High-resolution melting analysis of single nucleotide polymorphisms. Methods Mol. Biol. 1310, 5−27.

Cao, W., Bean, B., Corey, S., et al., 2015. Automated microfluidic platform for serial polymerase chain reaction and high-resolution melting analysis. J. Lab. Autom. 21 (3).

Carter, N.P., 2007. Methods and strategies for analyzing copy number variation using DNA microarrays. Nat. Genet. 39 (Suppl. 7), S16−S21.

Chagne, D., 2015. Application of the high-resolution melting technique for gene mapping and SNP detection in plants. Methods Mol. Biol. 1245, 151−159.

Chakravorty, S., Aladegbami, B., Burday, M., et al., 2010. Rapid universal identification of bacterial pathogens from clinical cultures by using a novel sloppy molecular beacon melting temperature signature technique. J. Clin. Microbiol. 48 (1), 258−267.

Chateigner-Boutin, A.L., Small, I., 2007. A rapid high-throughput method for the detection and quantification of RNA editing based on high-resolution melting of amplicons. Nucleic Acids Res. 35 (17), e114.

Chen, X.F., Li, T.R., Yang, H., et al., May 13, 2015. Detection of two drug-resistance mutants of the cytomegalovirus by high-resolution melting analysis. J. Clin. Lab. Anal. http://dx.doi.org/10.1002/jcla.21858 (Epub ahead of print).

Cheng, J.C., Huang, C.L., Lin, C.C., et al., 2006. Rapid detection and identification of clinically important bacteria by high-resolution melting analysis after broad-range ribosomal RNA real-time PCR. Clin. Chem. 52 (11), 1997–2004.

Cho, M.H., Ciulla, D., Klanderman, B.J., Raby, B.A., Silverman, E.K., 2008. High-resolution melting curve analysis of genomic and whole-genome amplified DNA. Clin. Chem. 54 (12), 2055–2058.

Crockett, A.O., Wittwer, C.T., 2001. Fluorescein-labeled oligonucleotides for real-time PCR: using the inherent quenching of deoxyguanosine nucleotides. Anal. Biochem. 290 (1), 89–97.

Cui, H., Dhroso, A., Johnson, N., Korkin, D., 2015. The variation game: cracking complex genetic disorders with NGS and omics data. Methods 79–80, 18–31.

Dahl, C., Guldberg, P., 2007. High-resolution melting for accurate assessment of DNA methylation. Clin. Chem. 53 (11), 1877–1878.

Dames, S., Margraf, R.L., Pattison, D.C., Wittwer, C.T., Voelkerding, K.V., 2007. Characterization of aberrant melting peaks in unlabeled probe assays. J. Mol. Diagn. 9 (3), 290–296.

Dimauro, S., Davidzon, G., 2005. Mitochondrial DNA and disease. Ann. Med. 37 (3), 222–232.

Do, H., Dobrovic, A., 2009. Limited copy number-high resolution melting (LCN-HRM) enables the detection and identification by sequencing of low level mutations in cancer biopsies. Mol. Cancer 8, 82.

Dobrovic, A., 2015. High-resolution assessment of copy number variation. Clin. Chem. 61 (5), 684–685.

Dobrowolski, S.F., Ellingson, C., Coyne, T., et al., 2007a. Mutations in the phenylalanine hydroxylase gene identified in 95 patients with phenylketonuria using novel systems of mutation scanning and specific genotyping based upon thermal melt profiles. Mol. Genet. Metab. 91 (3), 218–227.

Dobrowolski, S.F., Ellingson, C.E., Caldovic, L., Tuchman, M., 2007b. Streamlined assessment of gene variants by high resolution melt profiling utilizing the ornithine transcarbamylase gene as a model system. Hum. Mutat. 28 (11), 1133–1140.

Dobrowolski, S.F., Hendrickx, A.T., van den Bosch, B.J., et al., 2009. Identifying sequence variants in the human mitochondrial genome using high-resolution melt (HRM) profiling. Hum. Mutat. 30 (6), 891–898.

Dobrowolski, S.F., Pham, H.T., Downes, F.P., Prior, T.W., Naylor, E.W., Swoboda, K.J., 2012. Newborn screening for spinal muscular atrophy by calibrated short-amplicon melt profiling. Clin. Chem. 58 (6), 1033–1039.

Dufresne, S.D., Belloni, D.R., Wells, W.A., Tsongalis, G.J., 2006. BRCA1 and BRCA2 mutation screening using SmartCycler II high-resolution melt curve analysis. Arch. Pathol. Lab. Med. 130 (2), 185–187.

Dujols, V.E., Kusukawa, N., McKinney, J.T., Dobrowolski, S.F., Wittwer, C.T., 2006. High-resolution melting analysis for scanning and genotyping. In: Dorak, M.T. (Ed.), Real-Time PCR. Garland Science, New York, pp. 157–171.

Duyvejonck, H., Cools, P., Decruyenaere, J., et al., 2015. Validation of high resolution melting analysis (HRM) of the amplified ITS2 region for the detection and identification of yeasts from clinical samples: comparison with culture and MALDI-TOF based identification. PLoS One 10 (8), e0132149.

Dwight, Z., Palais, R., Wittwer, C.T., 2011. uMELT: prediction of high-resolution melting curves and dynamic melting profiles of PCR products in a rich web application. Bioinformatics 27 (7), 1019–1020.

Dwight, Z.L., Palais, R., Wittwer, C.T., 2012. uAnalyze: web-based high-resolution DNA melting analysis with comparison to thermodynamic predictions. IEEE ACM Trans. Comput. Biol. Bioinform. 9 (6), 1805–1811.

Dwight, Z.L., Palais, R., Kent, J., Wittwer, C.T., 2014. Heterozygote PCR product melting curve prediction. Hum. Mutat. 35 (3), 278–282.

Ebili, H.O., Ilyas, M., 2015. Cancer mutation screening: comparison of high-resolution melt analysis between two platforms. Ecancermedicalscience 9, 522.

Enache, E.L., Sin, A., Bancu, L., et al., 2015. Duplex high-resolution melting assay for the simultaneous genotyping of *IL28B* rs12979860 and *PNPLA3* rs738409 polymorphisms in chronic hepatitis C patients. Int. J. Mol. Sci. 16 (9), 22223–22242.

Er, T.K., Chang, J.G., 2012. High-resolution melting: applications in genetic disorders. Clin. Chim. Acta 414, 197–201.

Er, T.K., Kan, T.M., Su, Y.F., et al., 2012. High-resolution melting (HRM) analysis as a feasible method for detecting spinal muscular atrophy via dried blood spots. Clin. Chim. Acta 413 (21–22), 1781–1785.

Erali, M., Wittwer, C.T., 2010. High resolution melting analysis for gene scanning. Methods 50 (4), 250–261.

Erali, M., Voelkerding, K.V., Wittwer, C.T., 2008a. High resolution melting applications for clinical laboratory medicine. Exp. Mol. Pathol. 85 (1), 50–58.

Erali, M., Palais, R., Wittwer, C.T., 2008b. SNP genotyping by unlabeled probe melting analysis. In: Seitz, O., Marx, A. (Eds.), Molecular Beacons: Signalling Nucleic Acid Probes, Methods and Protocols. Humana Press, Totowa, pp. 199–206.

Esterhuizen, A.I., Wilmshurst, J.M., Goliath, R.G., Greenberg, L.J., 2014. Duchenne muscular dystrophy: high-resolution melting curve analysis as an affordable diagnostic mutation scanning tool in a South African cohort. S. Afr. Med. J. 104 (11), 779–784.

Farrar, J.S., Palais, R.A., Wittwer, C.T., 2011. Snapback primer genotyping of the Gilbert syndrome UGT1A1 (TA)(n) promoter polymorphism by high-resolution melting. Clin. Chem. 57 (9), 1303–1310.

Farrar, J.S., Witter, C.T., 2014. Extreme PCR: Efficient and Specific DNA amplification in 15-60 seconds. Clin. Chem. 61 (1), 145–153.

Fortini, D., Ciammaruconi, A., De Santis, R., et al., 2007. Optimization of high-resolution melting analysis for low-cost and rapid screening of allelic variants of Bacillus anthracis by multiple-locus variable-number tandem repeat analysis. Clin. Chem. 53 (7), 1377–1380.

Gelaye, E., Lamien, C.E., Silber, R., Tuppurainen, E.S., Grabherr, R., Diallo, A., 2013. Development of a cost-effective method for capripoxvirus genotyping using snapback primer and dsDNA intercalating dye. PLoS One 8 (10), e75971.

Gerini, C., Dal Canto, M., Porfirio, B., 2014. High resolution melting analysis of deletion/insertion polymorphisms: a new method for the detection and quantification of mixed chimerism in allogeneic stem cell transplantation. Mol. Cell Probe 28 (1), 19–24.

Germer, S., Higuchi, R., 1999. Single-tube genotyping without oligonucleotide probes. Genome Res. 9 (1), 72–78.

Giglio, S., Monis, P.T., Saint, C.P., 2003. Demonstration of preferential binding of SYBR Green I to specific DNA fragments in real-time multiplex PCR. Nucleic Acids Res. 31 (22), e136.

Gorniak, P., Ejduk, A., Borg, K., et al., 2015. Comparison of high-resolution melting analysis with direct sequencing for the detection of recurrent mutations in DNA methyltransferase 3A and isocitrate dehydrogenase 1 and 2 genes in acute myeloid leukemia patients. Eur. J. Haematol.

Graham, R., Liew, M., Meadows, C., Lyon, E., Wittwer, C.T., 2005. Distinguishing different DNA heterozygotes by high-resolution melting. Clin. Chem. 51 (7), 1295–1298.

Gundry, C.N., Vandersteen, J.G., Reed, G.H., Pryor, R.J., Chen, J., Wittwer, C.T., 2003. Amplicon melting analysis with labeled primers: a closed-tube method for differentiating homozygotes and heterozygotes. Clin. Chem. 49 (3), 396−406.

Gundry, C.N., Dobrowolski, S.F., Martin, Y.R., et al., 2008. Base-pair neutral homozygotes can be discriminated by calibrated high-resolution melting of small amplicons. Nucleic Acids Res. 36 (10), 3401−3408.

Guo, Q., Xiao, L., Zhou, Y., 2012. Rapid diagnosis of aneuploidy by high-resolution melting analysis of segmental duplications. Clin. Chem. 58 (6), 1019−1025.

Herrmann, M.G., Durtschi, J.D., Bromley, L.K., Wittwer, C.T., Voelkerding, K.V., 2006. Amplicon DNA melting analysis for mutation scanning and genotyping: cross-platform comparison of instruments and dyes. Clin. Chem. 52 (3), 494−503.

Herrmann, M.G., Durtschi, J.D., Bromley, L.K., Wittwer, C.T., Voelkerding, K.V., 2007a. Instrument comparison for heterozygote scanning of single and double heterozygotes: a correction and extension of Herrmann et al., Clin Chem 2006;52:494−503. Clin. Chem. 53 (1), 150−152.

Herrmann, M.G., Durtschi, J.D., Wittwer, C.T., Voelkerding, K.V., 2007b. Expanded instrument comparison of amplicon DNA melting analysis for mutation scanning and genotyping. Clin. Chem. 53 (8), 1544−1548.

Hinrichs, J.W., van Blokland, W.T., Moons, M.J., et al., 2015. Comparison of next-generation sequencing and mutation-specific platforms in clinical practice. Am. J. Clin. Pathol. 143 (4), 573−578.

Hoek, K.G., Gey van Pittius, N.C., Moolman-Smook, H., et al., 2008. Fluorometric assay for testing rifampin susceptibility of *Mycobacterium tuberculosis* complex. J. Clin. Microbiol. 46 (4), 1369−1373.

Holysz, M., Bialas, K., Migdalski, P., Kmieciak, D., Trzeciak, W.H., 2015. Application of a high-resolution melting technique for the rapid detection of partial replacement of HCV-1b by HCV-1a after PEG-IFN alpha/RBV therapy. J. Appl. Genet. 56 (2), 271−275.

Hondow, H.L., Fox, S.B., Mitchell, G., et al., 2011. A high-throughput protocol for mutation scanning of the BRCA1 and BRCA2 genes. BMC Cancer 11, 265.

Hoskins, J.M., Goldberg, R.M., Qu, P., Ibrahim, J.G., McLeod, H.L., 2007. UGT1A1*28 genotype and irinotecan-induced neutropenia: dose matters. J. Natl. Cancer Inst. 99 (17), 1290−1295.

Huang, K.T., Mikeska, T., Li, J., et al., 2015. Assessment of DNA methylation profiling and copy number variation as indications of clonal relationship in ipsilateral and contralateral breast cancers to distinguish recurrent breast cancer from a second primary tumour. BMC Cancer 15 (1), 669.

Innocenti, F., Undevia, S.D., Iyer, L., et al., 2004. Genetic variants in the UDP-glucuronosyltransferase 1A1 gene predict the risk of severe neutropenia of irinotecan. J. Clin. Oncol. 22 (8), 1382−1388.

Janukonyte, J., Vestergaard, E.M., Ladefoged, S.A., Nissen, P.H., 2010. High-resolution melting analysis using unlabeled probe and amplicon scanning simultaneously detects several lactase persistence variants. Scand. J. Clin. Lab. Inv. 70 (8), 535−540.

Jeffery, N., Gasser, R.B., Steer, P.A., Noormohammadi, A.H., 2007. Classification of Mycoplasma synoviae strains using single-strand conformation polymorphism and high-resolution melting-curve analysis of the vlhA gene single-copy region. Microbiology 153 (Pt 8), 2679−2688.

Jeng, K., Gaydos, C.A., Blyn, L.B., et al., 2012. Comparative analysis of two broad-range PCR assays for pathogen detection in positive-blood-culture bottles: PCR-high-resolution melting analysis versus PCR-mass spectrometry. J. Clin. Microbiol. 50 (10), 3287−3292.

Kristensen, L.S., Mikeska, T., Krypuy, M., Dobrovic, A., 2008. Sensitive melting analysis after real time- methylation specific PCR (SMART-MSP): high-throughput and probe-free quantitative DNA methylation detection. Nucleic Acids Res. 36 (7), e42.

Kristensen, L.S., Andersen, G.B., Hager, H., Hansen, L.L., 2012. Competitive amplification of differentially melting amplicons (CADMA) enables sensitive and direct detection of all mutation types by high-resolution melting analysis. Hum. Mutat. 33 (1), 264−271.

Lay, M.J., Wittwer, C.T., 1997. Real-time fluorescence genotyping of factor V Leiden during rapid-cycle PCR. Clin. Chem. 43 (12), 2262−2267.

Lee, T.H., Wu, T.S., Tseng, C.P., Qiu, J.T., 2012. High-resolution melting molecular signatures for rapid identification of human papillomavirus genotypes. PLoS One 7 (8), e42051.

Lehmensiek, A., Sutherland, M.W., McNamara, R.B., 2008. The use of high resolution melting (HRM) to map single nucleotide polymorphism markers linked to a covered smut resistance gene in barley. Theor. Appl. Genet. 117 (5), 721−728.

Li, M., Zhou, L., Palais, R.A., Wittwer, C.T., 2014. Genotyping accuracy of high-resolution DNA melting instruments. Clin. Chem. 60 (6), 864−872.

Liew, M., Pryor, R., Palais, R., et al., 2004. Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. Clin. Chem. 50 (7), 1156−1164.

Lin, J.H., Tseng, C.P., Chen, Y.J., et al., 2008. Rapid differentiation of influenza A virus subtypes and genetic screening for virus variants by high-resolution melting analysis. J. Clin. Microbiol. 46 (3), 1090−1097.

Lipsky, R.H., Mazzanti, C.M., Rudolph, J.G., et al., 2001. DNA melting analysis for detection of single nucleotide polymorphisms. Clin. Chem. 47 (4), 635−644.

Mackay, J.F., Wright, C.D., Bonfiglioli, R.G., 2008. A new approach to varietal identification in plants by microsatellite high resolution melting analysis: application to the verification of grapevine and olive cultivars. Plant Methods 4, 8.

Marcy, Y., Cousin, P.Y., Rattier, M., et al., 2008. Innovative integrated system for real-time measurement of hybridization and melting on standard format microarrays. Biotechniques 44 (7), 913−920.

Margraf, R.L., Mao, R., Wittwer, C.T., 2006a. Masking selected sequence variation by incorporating mismatches into melting analysis probes. Hum. Mutat. 27 (3), 269−278.

Margraf, R.L., Mao, R., Highsmith, W.E., Holtegaard, L.M., Wittwer, C.T., 2006b. Mutation scanning of the RET protooncogene using high-resolution melting analysis. Clin. Chem. 52 (1), 138−141.

Margraf, R.L., Mao, R., Highsmith, W.E., Holtegaard, L.M., Wittwer, C.T., 2007. RET proto-oncogene genotyping using unlabeled probes, the masking technique, and amplicon high-resolution melting analysis. J. Mol. Diagn. 9 (2), 184−196.

Mayerhofer, B., Stoger, A., Pietzka, A.T., et al., 2015. Improved protocol for rapid identification of certain spa types using high resolution melting curve analysis. PLoS One 10 (3), e0116713.

Mazi, W., Sangal, V., Sandstrom, G., Saeed, A., Yu, J., 2015. Evaluation of spa-typing of methicillin-resistant *Staphylococcus aureus* using high-resolution melting analysis. Int. J. Infect. Dis. 38, 125−128.

Mergny, J.L., Lacroix, L., 2003. Analysis of thermal melting curves. Oligonucleotides 13 (6), 515−537.

Meuzelaar, L., Hopkins, K., Liebana, E., Brookes, A., 2007. DNA diagnostics by surface-bound melt-curve reactions. J. Mol. Diagn. 9 (1), 30−41.

Minucci, A., Concolino, P., Giardina, B., Zuppi, C., Capoluongo, E., 2010. Rapid UGT1A1 (TA)(n) genotyping by high resolution melting curve analysis for Gilbert's syndrome diagnosis. Clin. Chim. Acta 411 (3−4), 246−249.

Mongelli, G., Bongiorno, D., Agosta, M., Benvenuto, S., Stefani, S., Campanile, F., 2015. High Resolution Melting-Typing (HRMT) of methicillin-resistant *Staphylococcus aureus* (MRSA): the new frontier to replace multi-locus sequence typing (MLST) for epidemiological surveillance studies. J. Microbiol. Methods 117, 136−138.

Montgomery, J., Wittwer, C.T., Kent, J.O., Zhou, L., 2007. Scanning the cystic fibrosis transmembrane conductance regulator gene using high-resolution DNA melting analysis. Clin. Chem. 53 (11), 1891−1898.

Montgomery, J.L., Sanford, L.N., Wittwer, C.T., 2010. High-resolution DNA melting analysis in clinical research and diagnostics. Expert Rev. Mol. Diagn. 10 (2), 219−240.

Morikawa, S., Harahap, I.S., Kaszynski, R.H., et al., 2011. Diagnosis of spinal muscular atrophy via high-resolution melting analysis symmetric polymerase chain reaction without probe: a screening evaluation for SMN1 deletions and intragenic mutations. Genet. Test. Mol. Biomark. 15 (10), 677−684.

Nguyen, Q., McKinney, J., Johnson, D.J., Roberts, K.A., Hardy, W.R., 2012. STR melting curve analysis as a genetic screening tool for crime scene samples. J. Forensic Sci. 57 (4), 887−899.

Nicklas, J.A., Noreault-Conti, T., Buel, E., 2012. Development of a fast, simple profiling method for sample screening using high resolution melting (HRM) of STRs. J. Forensic Sci. 57 (2), 478−488.

Odell, I.D., Cloud, J.L., Seipp, M., Wittwer, C.T., 2005. Rapid species identification within the Mycobacterium chelonae-abscessus group by high-resolution melting analysis of hsp65 PCR products. Am. J. Clin. Pathol. 123 (1), 96−101.

Ono, S., Domschke, K., Deckert, J., 2015. Genomic structural variation in affective, anxiety, and stress-related disorders. J. Neural Transm. 122 (1), 69−78.

Out, A.A., van Minderhout, I.J., van der Stoep, N., et al., 2015. High-resolution melting (HRM) re-analysis of a polyposis patients cohort reveals previously undetected heterozygous and mosaic APC gene mutations. Fam. Cancer 14 (2), 247−257.

Palais, R.A., Liew, M.A., Wittwer, C.T., 2005. Quantitative heteroduplex analysis for single nucleotide polymorphism genotyping. Anal. Biochem. 346 (1), 167−175.

Pang, Y., Liu, G., Wang, Y., Zheng, S., Zhao, Y.L., 2013. Combining COLD-PCR and high-resolution melt analysis for rapid detection of low-level, rifampin-resistant mutations in *Mycobacterium tuberculosis*. J. Microbiol. Methods 93 (1), 32−36.

Perera, P.K., Gasser, R.B., Pulford, D.J., et al., 2015. Comparison of the performance of three PCR assays for the detection and differentiation of Theileria orientalis genotypes. Parasite Vector 8, 192.

Petropoulou, M., Poula, A., Traeger-Synodinos, J., et al., 2015. Screening non-deletion alpha-thalassaemia mutations in the HBA1 and HBA2 genes by high-resolution melting analysis. Clin. Chem. Lab. Med. 53 (12), 1951−1959.

Poritz, M.A., Blaschke, A.J., Byington, C.L., et al., 2011. FilmArray, an automated nested multiplex PCR system for multi-pathogen detection: development and application to respiratory tract infection. PLoS One 6, e26047.

Pornprasert, S., Phusua, A., Suanta, S., Saetung, R., Sanguansermsri, T., 2008. Detection of alpha-thalassemia-1 Southeast Asian type using real-time gap-PCR with SYBR Green1 and high resolution melting analysis. Eur. J. Haematol. 80 (6), 510−514.

Price, E.P., Smith, H., Huygens, F., Giffard, P.M., 2007. High-resolution DNA melt curve analysis of the clustered, regularly interspaced short-palindromic-repeat locus of Campylobacter jejuni. Appl. Environ. Microbiol. 73 (10), 3431−3436.

Radvanszky, J., Surovy, M., Nagyova, E., Minarik, G., Kadasi, L., 2015. Comparison of different DNA binding fluorescent dyes for applications of high-resolution melting analysis. Clin. Biochem. 48 (9), 609−616.

Rajaei, M., Saadat, I., Saadat, M., 2012. Introducing a novel allele for the polymorphism of variable number of tandem repeats in the promoter region of XRCC5. Biochem. Biophys. Res. Commun. 427 (3), 503−505.

Rasmussen, J.P., Saint, C.P., Monis, P.T., 2007. Use of DNA melting simulation software for in silico diagnostic assay design: targeting regions with complex melting curves and confirmation by real-time PCR using intercalating dyes. BMC Bioinform. 8 (107), 107.

Reed, G.H., Wittwer, C.T., 2004. Sensitivity and specificity of single-nucleotide polymorphism scanning by high-resolution melting analysis. Clin. Chem. 50 (10), 1748−1754.

Reed, G.H., Kent, J.O., Wittwer, C.T., 2007. High-resolution DNA melting analysis for simple and efficient molecular diagnostics. Pharmacogenomics 8 (6), 597−608.

Reja, V., Kwok, A., Stone, G., et al., 2010. ScreenClust: advanced statistical software for supervised and unsupervised high resolution melting (HRM) analysis. Methods 50 (4), S10−S14.

Riordan, J.R., Rommens, J.M., Kerem, B., et al., 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 245 (4922), 1066−1073.

Ririe, K.M., Rasmussen, R.P., Wittwer, C.T., 1997. Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. Anal. Biochem. 245 (2), 154−160.

Ruskova, L., Raclavsky, V., 2011. The potential of high resolution melting analysis (HRMA) to streamline, facilitate and enrich routine diagnostics in medical microbiology. Biomed. Pap. 155 (3), 239−252.

Sanford, L.N., Wittwer, C.T., 2013. Monitoring temperature with fluorescence during real-time PCR and melting analysis. Anal. Biochem. 434 (1), 26−33.

Sanford, L.N., Wittwer, C.T., 2014. Fluorescence-based temperature control for polymerase chain reaction. Anal. Biochem. 448, 75−81.

Seipp, M.T., Durtschi, J.D., Liew, M.A., et al., 2007. Unlabeled oligonucleotides as internal temperature controls for genotyping by amplicon melting. J. Mol. Diagn. 9 (3), 284−289.

Seipp, M.T., Pattison, D., Durtschi, J.D., Jama, M., Voelkerding, K.V., Wittwer, C.T., 2008. Quadruplex genotyping of F5, F2, and MTHFR variants in a single closed tube by high-resolution amplicon melting. Clin. Chem. 54 (1), 108−115.

Smith, G.D., Zhou, L., Rowe, L.R., et al., 2011. Allele-specific PCR with competitive probe blocking for sensitive and specific detection of BRAF V600E in thyroid fine-needle aspiration specimens. Acta Cytol. 55 (6), 576−583.

Snell, C., Krypuy, M., Wong, E.M., kConFab Investigators, Loughrey, M.B., Dobrovic, A., 2008. BRCA1 promoter methylation in peripheral blood DNA of mutation negative familial breast cancer patients with a BRCA1 tumour phenotype. Breast Cancer Res. 10 (1), R12.

Song, C., Castellanos-Rizaldos, E., Bejar, R., Ebert, B.L., Makrigiorgos, G.M., 2015. DMSO increases mutation scanning detection sensitivity of high-resolution melting in clinical samples. Clin. Chem.

Spitzwieser, M., Holzweber, E., Pfeiler, G., Hacker, S., Cichna-Markl, M., 2015. Applicability of HIN-1, MGMT and RASSF1A promoter methylation as biomarkers for detecting field cancerization in breast cancer. Breast Cancer Res. 17 (1), 125.

Stephens, A.J., Inman-Bamber, J., Giffard, P.M., Huygens, F., 2008. High-resolution melting analysis of the spa repeat region of *Staphylococcus aureus*. Clin. Chem. 54 (2), 432−436.

Sumer Celebi, H., Ozdag, H., 2014. Comparison of high-resolution melting analysis to denaturing high performance liquid chromatography in the detection of point mutations in MEFV, F5, and F2 genes. Turkish J. Med. Sci. 44 (5), 713−719.

Sun, H., Yang, Y., Yang, L., et al., 2014. Snapback primer mediated clamping PCR for detection of EGFR and KRAS mutations in NSCLC patients by high resolution melting analysis. Biomed. Res. Int. 2014, 407537.

Sundberg, S.O., Wittwer, C.T., Howell, R.M., et al., 2014. Microfluidic genotyping by rapid serial PCR and high-speed melting analysis. Clin. Chem. 60 (10), 1306−1313.

Taylor, C.F., 2009. Mutation scanning using high-resolution melting. Biochem. Soc. Trans. 37 (Pt 2), 433−437.

Thomas, V., Mazard, B., Garcia, C., Lacan, P., Gagnieu, M.C., Joly, P., 2013. UGT1A1 (TA)n genotyping in sickle-cell disease: high resolution melting (HRM) curve analysis or direct sequencing, what is the best way? Clin. Chim. Acta 424, 258−260.

Thomas, H.R., Percival, S.M., Yoder, B.K., Parant, J.M., 2014. High-throughput genome editing and phenotyping facilitated by high resolution melting curve analysis. PLoS One 9 (12), e114632.

Toi, C.S., Dwyer, D.E., 2008. Differentiation between vaccine and wild-type varicella-zoster virus genotypes by high-resolution melt analysis of single nucleotide polymorphisms. J. Clin. Virol. 43 (1), 18−24.

Tong, S.Y., Giffard, P.M., 2012. Microbiological applications of high-resolution melting analysis. J. Clin. Microbiol. 50 (11), 3418−3421.

Tsai, T.S., Huang, M.Y., Chang, Y.T., et al., 2011. High-resolution melting analysis is a more effective approach for screening TSC genes mutations. Genet. Test. Mol. Biomark. 15 (6), 415−421.

Tsai, S.Y., Er, T.K., Lin, C.W., Chang, J.G., 2014. Comprehensive analysis of UGT1A1 polymorphisms through high-resolution melting analysis and DNA sequencing. Clin. Lab. 60 (6), 1015−1026.

van den Oever, J.M., van Minderhout, I.J., Harteveld, C.L., et al., 2015. A novel targeted approach for noninvasive detection of paternally inherited mutations in maternal plasma. J. Mol. Diagn. 17 (5), 590−596.

Vandersteen, J.G., Bayrak-Toydemir, P., Palais, R.A., Wittwer, C.T., 2007. Identifying common genetic variants by high-resolution melting. Clin. Chem. 53 (7), 1191−1198.

Vossen, R.H., Aten, E., Roos, A., den Dunnen, J.T., 2009. High-resolution melting analysis (HRMA): more than just sequence variant screening. Hum. Mutat. 30 (6), 860−866.

Vossen, R.H., van Duijn, M., Daha, M.R., den Dunnen, J.T., Roos, A., 2010. High-throughput genotyping of mannose-binding lectin variants using high-resolution DNA-melting analysis. Hum. Mutat. 31 (4), E1286−E1293.

Wang, K.C., Chang, C.C., Chang, Y.F., Wang, S.H., Chiang, C.K., Tsai, C.P., 2015a. Evaluation and characterization of a high-resolution melting analysis kit for rapid carrier-screening test of spinal muscular atrophy. J. Neurogenet. 1−4.

Wang, K., Mei, D.Y., Liu, Q.N., et al., 2015b. Research of methods to detect genomic mutations induced by CRISPR/Cas systems. J. Biotechnol. 214, 128−132.

White, H.E., Hall, V.J., Cross, N.C., 2007. Methylation-sensitive high-resolution melting-curve analysis of the SNRPN gene as a diagnostic screen for Prader-Willi and Angelman syndromes. Clin. Chem. 53 (11), 1960−1962.

Winkel, B.G., Hollegaard, M.V., Olesen, M.S., et al., 2011. Whole-genome amplified DNA from stored dried blood spots is reliable in high resolution melting curve and sequencing analysis. BMC Med. Genet. 12, 22.

Wittwer, C.T., 2009. High-resolution DNA melting analysis: advancements and limitations. Hum. Mutat. 30 (6), 857−859.

Wittwer, C.T., Kusukawa, N., 2011. Real-time PCR and melting analysis. In: Persing, D.H., Tenover, F.C., Hayden, R., Nolte, F., Tang, Y.W., Belkum, A.V. (Eds.), Molecular Microbiology: Diagnostic Principles and Practice, second ed. ASM Press, Washington, DC, pp. 63−82.

Wittwer, C.T., Reed, G.B., Ririe, K.M., 1994. Rapid cycle DNA amplification. In: Mullis, I.K., Ferre, F., Gibbs, R. (Eds.), The Polymerase Chain Reaction. Deerfield Beach, FL, pp. 174−181.

Wittwer, C.T., Ririe, K.M., Andrew, R.V., David, D.A., Gundry, R.A., Balis, U.J., 1997a. The LightCycler: a microvolume multisample fluorimeter with rapid temperature control. Biotechniques 22 (1), 176−181.

Wittwer, C.T., Herrmann, M.G., Moss, A.A., Rasmussen, R.P., 1997b. Continuous fluorescence monitoring of rapid cycle DNA amplification. Biotechniques 22 (1), 130−131, 134−138.

Wittwer, C.T., Reed, G.H., Gundry, C.N., Vandersteen, J.G., Pryor, R.J., 2003. High-resolution genotyping by amplicon melting analysis using LCGreen. Clin. Chem. 49 (6 Pt 1), 853−860.

Wojdacz, T.K., Dobrovic, A., 2007. Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation. Nucleic Acids Res.

Worm, J., Aggerholm, A., Guldberg, P., 2001. In-tube DNA methylation profiling by fluorescence melting curve analysis. Clin. Chem. 47 (7), 1183−1189.

Wu, Z., Zhang, Y., Zhang, X., et al., 2014a. A multiplex snapback primer system for the enrichment and detection of JAK2 V617F and MPL W515L/K mutations in Philadelphia-negative myeloproliferative neoplasms. Biomed. Res. Int. 2014, 458457.

Wu, J., Zhou, Y., Zhang, C.Y., et al., 2014b. Co-amplification at lower denaturation-temperature PCR combined with unlabeled-probe high-resolution melting to detect KRAS codon 12 and 13 mutations in plasma-circulating DNA of pancreatic adenocarcinoma cases. Asian Pac. J. Cancer Prevent. 15 (24), 10647−10652.

Yang, S., Ramachandran, P., Rothman, R., et al., 2009. Rapid identification of biothreat and other clinically relevant bacterial species by use of universal PCR coupled with high-resolution melting analysis. J. Clin. Microbiol. 47 (7), 2252−2255.

Zhang, H., Zhang, X., Wang, J., Xian, J., Chen, X., Zhang, W., 2015a. Comparison of high-resolution melting analysis, Sanger sequencing and ARMS for KRAS mutation detection in metastatic colorectal cancer. Clin. Lab. 61 (3−4), 435−439.

Zhang, L., Zhao, J., Cui, G., Wang, H., Wang, D.W., 2015b. Genotyping on ALDH2: comparison of four different technologies. PLoS One 10 (3), e0122745.

Zhao, M., Wang, Q., Wang, Q., Jia, P., Zhao, Z., 2013. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. BMC Bioinform. 14 (Suppl. 11), S1.

Zhou, L., Myers, A.N., Vandersteen, J.G., Wang, L., Wittwer, C.T., 2004a. Closed-tube genotyping with unlabeled oligonucleotide probes and a saturating DNA dye. Clin. Chem. 50 (8), 1328−1335.

Zhou, L., Vandersteen, J., Wang, L., et al., 2004b. High-resolution DNA melting curve analysis to establish HLA genotypic identity. Tissue Antigens 64 (2), 156−164.

Zhou, L., Wang, L., Palais, R., Pryor, R., Wittwer, C.T., 2005. High-resolution DNA melting analysis for simultaneous mutation scanning and genotyping in solution. Clin. Chem. 51, 1770−1777.

Zhou, L., Errigo, R.J., Lu, H., Poritz, M.A., Seipp, M.T., Wittwer, C.T., 2008. Snapback primer genotyping with saturating DNA dye and melting analysis. Clin. Chem. 54 (10), 1648−1656.

Zhou, L., Palais, R.A., Smith, G.D., Anderson, D., Rowe, L.R., Wittwer, C.T., 2010. Enrichment and detection of rare alleles by means of snapback primers and rapid-cycle PCR. Clin. Chem. 56 (5), 814−822.

Zhou, L., Palais, R.A., Ye, F., Chen, J., Montgomery, J.L., Wittwer, C.T., 2013. Symmetric snapback primers for scanning and genotyping of the cystic fibrosis transmembrane conductance regulator gene. Clin. Chem. 59 (7), 1052−1061.

Zhou, L., Palais, R.A., Paxton, C.N., Geiersbach, K.B., Wittwer, C.T., 2015. Copy number assessment by competitive PCR with limiting deoxynucleotide triphosphates and high-resolution melting. Clin. Chem.

Zou, H., Taylor, W.R., Harrington, J.J., et al., 2009. High detection rates of colorectal neoplasia by stool DNA testing with a novel digital melt curve assay. Gastroenterology 136 (2), 459−470.

Chapter 7

Molecular Techniques for DNA Methylation Studies

J. Tost[1] and I.G. Gut[2]
[1]CEA—Institut de Génomique, Evry, France; [2]Barcelona Institute for Science and Technology, Barcelona, Spain

7.1 INTRODUCTION

All cells of a multicellular organism carry the same genetic code in their DNA sequence although cells display broad morphologic and functional diversity. The heterogeneity of cells is caused by the differential expression of genes. Epigenetics can be defined as the study of mitotically (and in some cases meiotically) heritable changes of a phenotype, such as the gene expression of specific cell types that do not result from changes in the genetic code (Waddington, 1942). Epigenetic regulation mediates the adaption to an environment ultimately contributing to the phenotype.

Epigenetic phenomena are mediated by a variety of molecular mechanisms including posttranscriptional histone modifications, histone variants, adenosine triphosphate (ATP)-dependent chromatin remodeling complexes, polycomb—trithorax protein complexes, small and other noncoding RNAs including small interfering RNAs and microRNAs (miRNAs), and DNA methylation (Tost, 2008). These diverse molecular mechanisms all have been found to be closely intertwined and stabilize each other to ensure the faithful propagation of an epigenetic state over time, especially through cell division. Because the description of all methods used for the analysis of the different kinds of epigenetic modifications is beyond the scope of this chapter, the authors will concentrate on DNA methylation as the best studied epigenetic alteration. As such, DNA methylation has great potential as a diagnostic and predictive tool as well as a therapeutic target. Nonetheless, the expression of long and small noncoding RNAs has also been shown to be altered in many diseases (Dey et al., 2014; Vicente et al., 2015). miRNA profiles detect not only early-stage cancer but also the determination of its tissue of origin when profiling circulating miRNAs (Rosenfeld et al., 2008; Schwarzenbach et al., 2014), and miRNA expression signatures have been identified that are associated with prognosis and response to therapy in different human cancers (Yu et al., 2008).

This chapter focuses on DNA methylation as a covalent DNA modification, but it should be kept in mind that transcription does not occur on unpacked DNA but rather in the context of chromatin, which critically influences the accessibility of the DNA to transcription factors and the DNA polymerase complexes. Chromatin modulations have a central role in shaping the epigenome and delineate a functional chromatin topology, which serves as the platform forming regulatory circuits in all cells. Open (euchromatin) and closed (heterochromatin) chromatin states are controlled by histone modifications, histone composition, and the ATP-dependent chromatin remodeling machinery in close cross-talk with the binding of a plethora of nonhistone proteins (Fig. 7.1). Positioning of nucleosomes, the degree of compaction (i.e., nucleosome spacing), and remodeling of chromatin have key roles in the coordination of the correct gene expression program. Positioning of nucleosomes depends on (among others) the underlying DNA sequence, ATP-dependent nucleosome remodelers, DNA-binding proteins, the RNA polymerase II transcription machinery, and their interactions (Struhl and Segal, 2013). As a result, the core enhancer, promoter, and terminator regions of genes are typically depleted of nucleosomes, whereas most of the genomic DNA is occupied. Consequently the analysis of chromatin accessibility and nucleosome positioning is essential for the understanding of transcriptional regulation and can be used for the analysis and identification of gene regulatory elements and their changes in disease, but is beyond the scope of this chapter.

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00007-9
Copyright © 2017 Elsevier Ltd. All rights reserved.

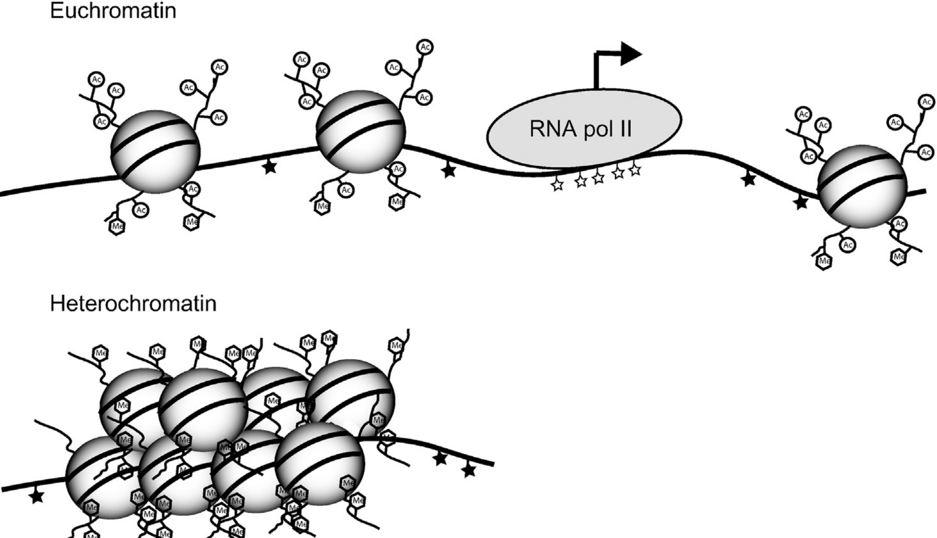


FIGURE 7.1 Simplified representation of euchromatin and heterochromatin. In euchromatin, RNA polymerase II (RNA pol II) can access the promoter CpG island (*open stars*), which corresponds to a nucleosome-free region and the region is permissive to transcription. Histones are hyperacetylated (*circles* with Ac) including lysine 9 of histone H3 (H3K9), and lysine 4 of H3 (H3K4) is methylated (*polygon* with Me). In contrast, the DNA and nucleosome structure is compacted in heterochromatin. H3K4 is demethylated; H3K9, H3K27, and H4K20 are methylated. DNA as well is hypermethylated and the region is transcriptionally silent. For reasons of clarity, only selected histone modifications associated with the chromatin state are shown.

DNA methylation is highly correlated with these chromatin modifications and enzymes that modify DNA and histones have been shown to interact directly and constitute links between local DNA methylation and regional chromatin structure (Geiman and Robertson, 2002).

Technology for the analysis of DNA methylation patterns have largely evolved over the years since the publication of the previous edition, and many technologies have been abandoned. For example, microarray-based technologies have mostly been replaced by sequencing-based approaches. Therefore, in this chapter we concentrate on assays that are currently widely used for genome-wide discovery as well as assays that can be employed in a clinical laboratory for the analysis of DNA methylation patterns as biomarkers.

7.1.1 The Biology of DNA Methylation

DNA methylation is the only genetically programmed DNA modification in mammals. This postreplication modification is found almost exclusively on the five position of the pyrimidine ring of cytosines in the context of the dinucleotide sequence CpG, with approximately 29 million CpGs present in the human genome (Fig. 7.2) (Rollins et al., 2006). 5-Methylcytosine accounts for approximately 1% of all bases, varying slightly in different tissue types; most CpG dinucleotides throughout mammalian genomes are methylated (75%). Other types of methylation, such as the methylation of cytosines in the context of CpNpG or CpA sequences, have been detected in mouse embryonic stem cells and plants but are generally rare in somatic mammalian or human tissues. Close to half of all CpGs are found in repetitive elements. The variety of epigenetic modifications in mammalian DNA has increased further with the discovery of 5-hydroxymethylcytosine (hmC) in mammalian neurons and embryonic stem cells, which is formed from 5-methylcytosines by a catalytic oxidation mediated by the TET proteins (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). This base modification has been found in nearly all mouse embryonic tissues, but with the exception of brain tissue and bone marrow, only low levels of hmC (0.2−0.05% of all cytosines) are detectable in adult tissue (Globisch et al., 2010; Ruzov et al., 2011). Its potential implication in disease has yet to be defined, but its high presence in embryonic, adult, or induced pluripotent stem cells suggest a biological role of this epigenetic

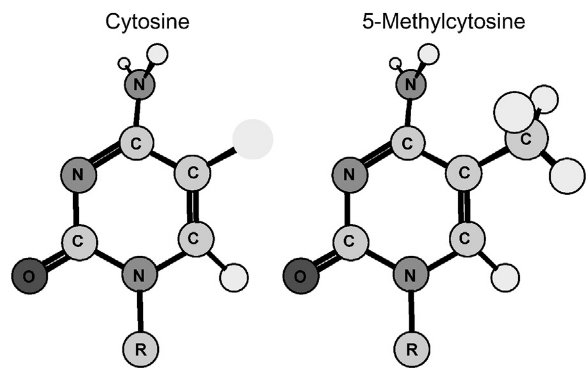


FIGURE 7.2 Chemical structure of cytosine and 5-methylcytosine. Cytosine is incorporated into the DNA using deoxycytidinetriphosphate as the building block and is methylated after its incorporation by DNA methyltransferases.

mark in pluripotency (Ruzov et al., 2011). 5-hmC has probably also a biochemical role as an intermediate in the process of DNA demethylation, where it undergoes oxidation to 5-formylcytosine and 5-carboxylcytosine by the TET oxygenases before being excised by a thymine-DNA glycosylase (He et al., 2011; Maiti and Drohat, 2011). However, although these modifications are certainly of great biological importance, their low prevalence (except for brain tissue) and the early stage of research into their distribution makes them currently less suitable for use as biomarkers. We will therefore focus in the current review on 5-methylcytosine methylation, referred to here as DNA methylation.

CpGs are underrepresented in the genome, probably because they are mutational hot spots, because the deamination of methylated CpGs to TpGs does not create unnatural bases and is recognized less efficiently by the DNA repair machinery. Despite this general trend, CpG-rich clusters approximately 1−4 kb in length (so-called CpG islands) are found in the promoter regions and first exons of many genes. CpGs in these islands representing about 7% of all CpGs are mostly not methylated, which corresponds to the maintenance of an open chromatin structure and a potentially active state of transcription (Antequera, 2003). There are around 30,000 CpG islands in the human genome. About three-quarters of transcription start sites and 88% of active promoters are associated with CpG-rich sequences and might be regulated by DNA methylation (Rollins et al., 2006). Methylated CpG islands are found on the inactive X chromosome in females, associated with genes with tissue-specific expression and in intragenic CpG islands. Evidence emphasizes the importance of the underlying DNA sequence for the establishment of DNA methylation patterns (Schubeler, 2015). Conversely, CpG islands facilitate alterations of the DNA conformation to Z-DNA, which has reduced affinity for nucleosomes. Of note, an unmethylated state of a CpG island does not necessarily correlate with the transcriptional activity of the gene, but rather with its potential for activation. On the other hand, the simple presence of methylation does not necessarily induce silencing of nearby genes. Only when a specific core region of the promoter becomes hypermethylated is the expression of the associated gene modified (Ushijima, 2005). DNA methylation often changes after modification of the chromatin structure and is used as a molecular mechanism to lock the gene permanently and thus heritably in its inactive state (Bird, 2002). Methylation can interfere with transcription in several ways. It can inhibit the binding of transcriptional activators with their cognate DNA recognition sequence such as Sp1 and Myc through sterical hindrance. Methyl-CpG-binding domain (MBD) proteins and the DNA methyltransferases themselves bind to methylated DNA and thereby prevent binding of potentially activating transcription factors (Defossez and Stancheva, 2011). These two protein families also recruit additional proteins with repressive function such as histone deacetylases and chromatin remodeling complexes to the methylated DNA to establish a repressive chromatin configuration creating a self-enforcing mechanism to silence genes.

7.1.2 Dynamics of DNA Methylation

During development and differentiation, the mammalian organism creates a number of cell type−specific differentially marked epigenomes whose identity is defined, among other things, by their respective DNA methylation patterns. Consequently, one genome contains approximately 180 different epigenomes that are currently mapped in large-scale efforts such as the Roadmap or the BLUEPRINT consortium (Adams et al., 2012; Roadmap Epigenomics Consortium et al., 2015). Cytosine methylation is essential for mammalian embryogenesis, which is characterized by two waves of genome-wide epigenetic reprogramming in the zygote and in the primordial germ cells (Reik et al., 2001). Incomplete reprogramming of these epigenetic patterns is one reason for the low success rate of somatic cell nuclear transfer, that is, the fusion of a somatic cell with an enucleated oocyte (Krishnakumar and Blelloch, 2013). Modifications in the environment during early development can lead to permanent changes in the patterns of epigenetic modifications. This modulation of epigenetic patterns in utero has resulted in the developmental origin of disease hypothesis, which postulates that the in utero environment can cause permanent changes to metabolic processes that directly affect postnatal phenotype, confers susceptibility to multifactorial disease at adult age, and may also be transmitted to subsequent generations (Hanson and Gluckman, 2014). Epigenetics also holds the promise to explain at least a part of the influences of the environment. Studies in monozygotic twins demonstrated that epigenetic differences in genetically identical humans (monozygotic twins) accumulate with age and different environments create different patterns of epigenetic modifications (Fraga et al., 2005). Monozygotic twins are nonetheless a powerful model for identifying the disease-relevant epigenetic changes for complex diseases because they avoid confounding effects of population-based studies owing to their identical genetic sequence and their shared early- and sometimes even late-life environmental exposure (Castillo-Fernandez et al., 2014; Tost et al., 2014). Chemical and environmental toxins have shown to induce changes to DNA methylation patterns without altering the genetic sequence and leading to epimutation-associated phenotypes (Green and Marsit, 2015). DNA methylation has multiple essential roles in mammals. It is critical for imprinting, that is, the asymmetric expression of either the maternal or paternal allele in a parent-of-origin−specific manner in somatic cells of the offspring (Reik and Walter, 2001; Sanli and Feil, 2015) is implicated in X chromosome inactivation, the random silencing of one of the two X chromosomes

in embryonic tissues of female mammals to achieve dosage compensation (Gendrel and Heard, 2014). It is required for the maintenance of genome integrity (Eden et al., 2003; Meng et al., 2015) and has a key role in transcriptional silencing of repetitive DNA sequences and endogenous transposons (Yoder et al., 1997; Crichton et al., 2014).

7.1.3 DNA Methylation and Disease

DNA methylation and chromatin structure are strikingly altered in many pathologic situations, particularly cancer. Although a number of genetic variations associated with disease susceptibility have been identified by genome-wide studies in cancer (Fachal and Dunning, 2015), neurodegenerative diseases (Chouraki and Seshadri, 2014), and metabolic disorders (Mohlke and Boehnke, 2015), they confer only small increases in risk and do not report on the onset of disease. Many of the susceptibility loci also do not associate with genes, but the strongest association is found in noncoding regions of the genome. Characterizing the epigenomic landscape in disease-relevant cell populations at the susceptibility loci might help to decipher the potential functional effect of these genetic variations (Tak and Farnham, 2015). Novel genome and epigenome engineering technologies now determine whether genetic and epigenetic changes are causal from observed phenotypic changes and will substantially further our understanding of the multifaceted roles of epigenetic modifications and their combinations in chromatin, which is important to the answer of many biomedical questions (Jurkowski et al., 2015; Thakore et al., 2015; Tost, 2016).

Furthermore, environmental factors undoubtedly have a large role in the actual occurrence of disease. The epigenome constitutes the memory of an organism to all of the stimuli and influences to which it has ever been exposed (Ladd-Acosta, 2015). Aberrant methylation patterns have been reported in various neurodevelopmental disorders and imprinting anomalies lead to disorders such as Prader–Willi, Angelman, and Beckwith-Wiedemann syndromes (Eggermann et al., 2015). DNA methylation patterns are globally disturbed in autoimmune diseases such as the lupus erythematosus (Absher et al., 2013), primary Sjögren syndrome (Miceli-Richard et al., 2015), and rheumatoid arthritis (Neidhart et al., 2000) and inflammatory diseases (Karatzas et al., 2014; Chandra et al., 2015). Epigenetic changes contribute to the susceptibility and development of many complex or multifactorial diseases (Ronn and Ling, 2015; Yu et al., 2015; Voelter-Mahlknecht, 2016). Epigenetic mechanisms are consistent with various non-Mendelian features of multifactorial diseases such as the relatively high degree of discordance in monozygotic twins. To underline the scope of epigenetic alterations in disease further, it is interesting that monogenetic diseases such as α-thalassemia, which were previously attributed solely to genetic alterations, can also be caused by epigenetic alterations at the same locus (Tufarelli et al., 2003) and the penetrance of genetic diseases can be modified by DNA methylation alterations (Jones et al., 2015).

The high frequency of DNA methylation changes in cancer and the potential diagnostic use of DNA methylation for diagnostics and prognostics have fueled much of the rapid catch-up of technologies for epigenetic analyses that we have seen in past years. We will devote most of the rest of the chapter to describing the applications of DNA methylation techniques in oncology.

7.1.3.1 DNA Methylation Changes in Cancer

Cancer is probably the best-studied disease with a strong epigenetic component (Jones and Baylin, 2007). In tumors a global loss of DNA methylation (hypomethylation) of the genome is observed (Feinberg and Vogelstein, 1983) and has been suggested to initiate and propagate oncogenesis by inducing chromosome instabilities and transcriptional activation of oncogenes and pro-metastatic genes (Chen et al., 1998; Ehrlich, 2002). The overall decrease in DNA methylation is accompanied by a region- and gene-specific increase in methylation (hypermethylation) of multiple CpG islands (Fig. 7.3) (Laird, 2005; Jones and Baylin, 2007). Hypermethylation of CpG islands in the promoter region of a tumor suppressor or otherwise cancer-related genes is often associated with transcriptional silencing of the associated gene. The number of gene-associated promoters that are known to become hypermethylated during carcinogenesis is rapidly growing. Genes of numerous pathways involved in signal transduction (*APC*), DNA repair (*MGMT*, *MLH1*, and *BRCA1*), detoxification (*GSTP1*), cell-cycle regulation (*p15*, *p16*, and *RB*), differentiation (*MYOD1*), angiogenesis (*THBS1*, *VHL*), and apoptosis (*Caspases*, *p14*, and *DAPK*) are often inappropriately inactivated by DNA methylation. To date, no single gene has been identified that is always methylated in a certain type of cancer. Both hypomethylation and hypermethylation are found in the same tumor, but the underlying mechanisms for both phenomena have not yet been elucidated. Furthermore, epigenetic modifications contribute to long-range, gene-silencing, large chromosomic regions (Frigola et al., 2006). Long-range epigenetic silencing seems to be a prevalent phenomenon during carcinogenesis, because a survey identified 28 regions of copy number–independent transcriptional deregulation in bladder cancer that are potentially regulated through epigenetic mechanisms (Stransky et al., 2006). Although the contribution of genetic factors to carcinogenesis such as the high-penetrance germ-line mutations in genes such as *BRCA1* and *p53* in familial cancers has long been recognized, it has become evident that epigenetic changes leading to transcriptional silencing of tumor suppressor genes

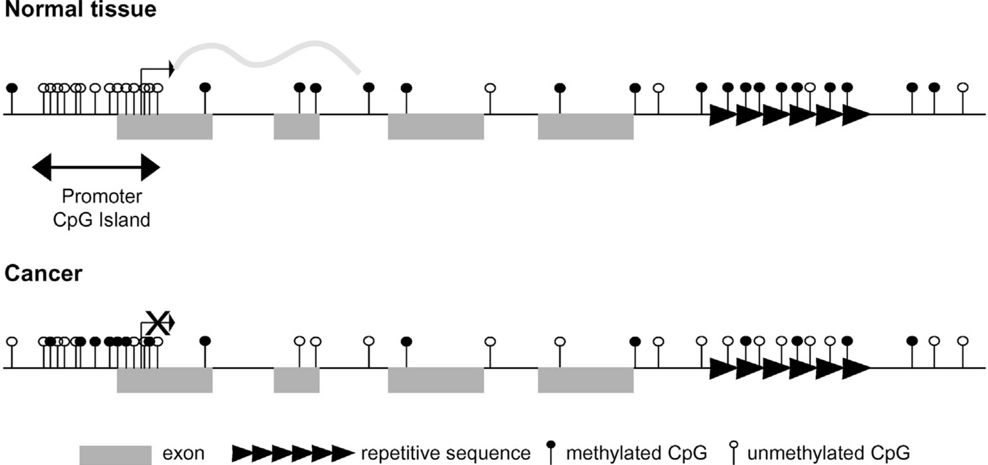


FIGURE 7.3 Distribution of DNA methylation in normal tissue and cancer. In the normal tissue, most promoter CpG islands are free of DNA methylation (*white circles*) even if the gene is not expressed. Repetitive elements as well as interspersed CpG dinucleotides are mostly methylated (*black circles*). In tumors, a global loss of DNA methylation (hypomethylation of the cancer genome) is observed whereas some promoter CpG islands become methylated in a tumor type–specific manner. Methylation patterns are dynamic and also change to a lesser extent during age and in response to environmental factors.

constitute an at least equally contributing mechanism. For example, microarray expression profiles of breast tumors with *BRCA1* mutations are similar to those of cases of sporadic breast cancer with *BRCA1* promoter hypermethylation. This demonstrates that disruption of *BRCA1* function by genetic or epigenetic pathways leads to the same perturbations (Hedenfalk et al., 2001). It has been extrapolated that aberrant promoter methylation is initiated at approximately 1% of all CpG islands and as much as 10% become methylated during the multistep process of tumorigenesis (Costello et al., 2000).

7.2 CLINICAL APPLICATIONS OF DNA METHYLATION ANALYSIS

Biomarkers capable of distinguishing diseased or malignant cells from normal ones must be specific, sensitive, and detectable in specimens obtained through minimally invasive procedures to be clinically applicable. Many biomarkers have been discovered on the level of protein, RNA, or DNA that fulfill these criteria. DNA-based molecular biomarkers are relatively easily transferred from a research laboratory setting into routine diagnostics owing to the amplifiable and stable nature of DNA. Methyl groups on cytosines are part of the covalent structure of the DNA. Once methylation is acquired, in most cases it is chemically and biologically stable over time whereas expression of mRNA and/or proteins can be modified by non–disease related environmental conditions and vary over the cell cycle. Most DNA methylation analysis methods determine the ratio of methylated and unmethylated CpGs and are independent of the total amount of starting material. It provides a binary and positive signal that can be detected independent of expression levels and is therefore easier to detect than negative signals such as loss of heterozygosity. DNA methylation can be analyzed by an increasing number of methods that are amenable to high-throughput, and quantitative assays eliminate the need for normalization. Compared with genetic alterations such as mutations, which are usually distributed throughout the gene, DNA methylation alterations are concentrated in a defined area, often the promoter, facilitating analysis. DNA methylation can be reliably analyzed in fresh-frozen and archived clinical specimens, but more important, also in biofluids that were either in direct contact with the disease (e.g. stool for colorectal cancer, urine for genitourinary cancer, or sputum or bronchoalveolar lavages or brushes for lung cancer) or in cell-free circulating DNA isolated from serum or plasma of patients (How-Kit et al., 2012).

Epigenetic changes occur at higher frequency compared with genetic changes and may be especially important in early-stage human neoplasia (Feinberg et al., 2006). Alterations of DNA methylation patterns often precede malignancy because extensive CpG island hypermethylation can be detected in benign polyps of the colon, in low- as well as high-grade tumors (Goelz et al., 1985; Costello et al., 2000). It has therefore been suggested that epigenetic lesions in normal tissue set the stage for neoplasia and might induce an addiction to altered signal transduction pathways leading to additional genetic changes (Baylin and Ohm, 2006). For example, DNA hypermethylation could be detected not only in dysplastic epithelium of patients with ulcerative colitis, a condition associated with an increased risk for the development of colon cancer, but in histologic normal epithelium (Issa et al., 2001). Aberrant DNA methylation of the tumor suppressor gene *CDKN2A*

in smokers was detectable up to 3 years before individuals were diagnosed with squamous cell lung carcinoma (Palmisano et al., 2000). Not only smokers but also never-smokers exposed to second-hand tobacco had a tendency toward *CDKN2A* hypermethylation (Scesnaite et al., 2012). Wielscher et al. (2015) developed a four-gene signature that potentially monitors people with pulmonary fibrosis and chronic obstructive lung disease who have a high risk of progressing to lung cancer, thus allowing early detection for malignant disease. Studies such as these suggest that by screening high-risk groups for aberrant DNA methylation it may be possible to identify subgroups of individuals who carry an increased risk for tumor development. Thus, treatment can be initiated at an earlier stage, gaining valuable time for therapy.

Besides early detection, the methylation status of CpG islands can be used to characterize and classify cancers. Distinct methylation profiles for various cancers have emerged in past years and have allowed the characterization and classification of these tumors based on CpG island methylation status (Eads et al., 2000b; Costello et al., 2000; Fernandez et al., 2012).

Prognostic markers assist in predicting the different risks of outcome for patients regardless of the type of treatment. Contrary to predictive markers, they may not be a target of the therapy. Biomarkers of highest clinical value would be able to evaluate the malignant potential of a tumor and identify patients at high risk for rapid tumor progression and/or tumor recurrence. Moreover, prognostic markers that can predict decreased survival time might help stratify different tumor subtypes in which additional adjuvant therapy or a combination of therapies might be required, thereby avoiding loss of valuable time owing to ineffective treatment. Many studies have investigated the potential of DNA methylation of some candidate genes as prognostic biomarkers in different types of cancers. Genome-wide methylation profiling in ovarian cancers defined tumor subgroups in which those with CpG island hypermethylation correlated with early disease recurrence after chemotherapy (Wei et al., 2002). Moreover, 112 methylated prognostic loci correlated with progression-free survival at an accuracy of 95% (Wei et al., 2006) and 20 methylated CpG island loci including *WNT* pathway genes were associated with progression of ovarian cancer (Dai et al., 2011). In acute myeloid leukemia, a genome-wide methylation experiment found a combination of 15 genes among 14,000 analyzed as potential biomarkers for overall survival (Figueroa et al., 2010), whereas 27,000 epigenotyping arrays demonstrated an association between low global DNA methylation and higher complete remission rate and *CDKN2B* methylation was associated with better overall and disease-free survival (Deneberg et al., 2010). The latter technology also led to the identification of 100 methylated loci associated with disease progression in breast cancer (Fackler et al., 2011) and six CpG signatures associated with overall survival in glioblastoma (Etcheverry et al., 2010). In colon cancer, a genome-wide genomic and epigenomic study identified hypermethylated biomarkers from extracellular remodeling pathways associated with poor survival (two markers) or poor survival (multiple markers) (Yi et al., 2011). These examples clearly demonstrate the power of genome-wide approaches to identify the best prognostic methylation-based biomarkers.

However, despite the large number of prognostic markers reported, many of the reported studies, especially those analyzing few selected candidate genes, lack sufficient power and few prognostic biomarkers have been confirmed in independent data sets. Evaluation is further complicated by the fact that if prognostic values are investigated in several studies, they are difficult to compare because of varying and/or inadequately described patient cohorts of often-limited sample size. Moreover, studies frequently use different assays, which give dichotomized or quantitative outputs, and which were further treated using a number of data transformation steps such as "arbitrarily" or "*p* value−driven"-defined threshold levels before the assessment of the prognostic value. A probably large reporting bias toward positive associations combined with a lack of randomization in the original observational study further complicates a clear assessment of their clinical value. It is therefore currently difficult to assess the true prognostic power of DNA methylation-based biomarkers, and adequately powered, well-designed studies with clear reporting guidelines will be important for the field of DNA methylation-based prognostic markers to move forward.

To date, clinical applications of DNA methylation have focused mainly on cancer-related biomarkers and little has been done on other complex diseases. The report on the detection of aberrantly methylated molecules in cell-free circulating DNA in lupus erythematosus opens the way to perform analyses similar to those performed in the field of cancer epigenomics as for other complex diseases (Chan et al., 2014).

Several DNA methylation-based biomarkers are commonly used in clinics, such as testing of the DNA methylation status of the *MGMT* promoter to predict the response to chemotherapy using alkylating agents in glioblastoma (Hegi et al., 2005) or of *MLH1* to diagnose Lynch syndrome (Newton et al., 2014). *MGMT* encodes a DNA repair protein that removes alkyl groups from the O6-position of guanines protecting DNA from damage. The absence of MGMT could thus promote tumorigenesis through an increased mutation rate. The hypermethylation of the *MGMT* promoter frequently occurs in a variety of cancers including colorectal; lung and head and neck cancers; lymphomas; and, of course, gliomas (Esteller et al., 1999; Jacinto and Esteller, 2007; von Deimling et al.,

2011). *MGMT* silencing in cancer results from CpG island promoter hypermethylation associated with repressive histone modifications such as H3K9 dimethylation (Nakagawachi et al., 2003). In high-grade gliomas, three important molecular biomarkers have been identified: 1p/19q co-deletion, *IDH1/2* mutation, and *MGMT* promoter methylation, all of which have prognostic values, whereas only *MGMT* promoter methylation carries predictive information (Tabatabai et al., 2010). The *MGMT* promoter methylation status contains both prognostic and predictive information, depending on the grade of the gliomas. Indeed, whereas *MGMT* promoter methylation prognoses longer progression-free and overall survival in all types of anaplastic gliomas independent of the type of chemotherapy (Brandes et al., 2006; van den Bent et al., 2009; Wick et al., 2009), it predicts favorable outcome in patients treated with alkylating agents such as temozolomide in glioblastomas (Tabatabai et al., 2010).

Furthermore, commercial products are available, such as the Epi *pro*Colon test, which analyzes methylation in the *SEPT9* gene (Church et al., 2014) for population-wide screening for colorectal cancer [approved by the Chinese Food and Drug Administration (FDA) in July 2015 and by the US FDA in April 2016] and the *Epi pro*Lung (*SHOX2*) test (Ilse et al., 2014). Panels with multiple genes can further improve the specificity and sensitivity of DNA methylation signatures, as demonstrated in a stool-based screening test for colorectal cancer (Imperiale et al., 2014) (Cologuard, Exact Sciences), in which DNA methylation alterations were combined with mutation detection, and which received the first US FDA approval for a DNA methylation-based diagnostic test in 2014.

7.3 METHODS FOR DNA METHYLATION ANALYSIS

For the analysis of DNA methylation, sensitive and quantitative methods are required to detect even subtle changes in the degree of methylation because biological samples often represent a heterogeneous mixture of different cells, especially tumor and nontumor cells from tissue biopsies. With the realization of the importance of epigenetic changes in development and disease, a variety of techniques for the study of DNA methylation has been developed over the past years (Fouse et al., 2010; Harris et al., 2010; How-Kit et al., 2012; Olkhov-Mitsel and Bapat, 2012; Lee et al., 2013; Sun et al., 2015). No single method has emerged as the reference standard technique unifying quantitative accuracy and high sensitivity, possibility for whole-genome analysis and the precise investigation of individual CpG positions. Thus, the choice of the method depends mainly on the desired application and the type, quality, and quantity of the available biological material. Essentially, methods can be classified into two categories: genome-wide and targeted. Several generations of genome-wide methods have been developed adapting to different analytical supports with increasing levels of resolution whereas targeted approaches vary in the function of their application.

7.3.1 The Methylation Content of a Sample

Methods for the analysis of global DNA methylation levels in a sample determine the overall 5-methylcytosine content or changes affecting the entire epigenome, respectively (Table 7.1). However, they give no information about the location or repartition of DNA methylation on the genome. 5-Methylcytosine can be differentiated from its unmethylated counterpart by the different mass or polarity of the two bases (cytosine and 5-methylcytosine), which can be used for chromatographic or mass spectrometric separation (Wiebers, 1976; Eick et al., 1983; Fraga et al., 2002). The 5-methylcytosine content is measured after hydrolysis to mononucleosides, and comparison with an internal standard enables quantification. Particularly since the discovery of the oxidative derivatives of 5-methylcytosine, the mass spectrometry detection has become popular, also owing to its exquisite sensitivity and the low amounts of required starting material (Globisch et al., 2010; Le et al., 2011).

Bacterial methyltransferases such as *Sss*I transfer a methyl group from the universal methyl donor *S*-adenosyl-L-methionine to unmethylated CpG positions. The methyl acceptor assay uses these enzymes and analyses the amount of incorporated radiolabeled methyl groups into a sample (Bestor and Ingram, 1983). The measured amount of radioactive label thus correlates inversely with the degree of its methylation before labeling. Similarly, the cytosine extension assay combines methylation-sensitive (MS) restriction digestion and single-nucleotide extension with radiolabeled or fluorescently labeled deoxycytidine triphosphate complementary to the guanine 5′ overhang created by the digestion (Pogribny et al., 1999; Bönsch et al., 2004). The Pyrosequencing™ Luminometric Methylation Assay is based on the differential digestion of a sample with an MS endonuclease or its methylation-insensitive isoschizomer and the successive dispensation of four nucleotides complementary to the overhang created by the endonucleases (Karimi et al., 2006). The Pyrosequencing™-based analysis of repetitive elements such as Alu and LINE1 has also been widely used as surrogate for the global DNA methylation level (Yang et al., 2004). Methods analyzing the total amount of 5-methylcytosine in a sample are used to analyze and follow global DNA methylation changes induced by demethylating pharmaceutical agents in patients with hematologic malignancies at various points of treatment

TABLE 7.1 Methods for Genome-Wide Identification of Differentially Methylated Genes

Method	Principle of Differentiation	Method	Enrichment of Methylated/ Unmethylated Fragments	Detection Platform	Comment	References
Methylation reversal	Reactivation of epigenetically silenced genes by treatment with methylation inhibitors	RNA expression analysis before and after treatment with demethylating agents	–	Complementary DNA/expression microarrays/RNA-sequencing	Identifies epigenetic modifications associated with transcriptional changes, potential confounding effects of demethylating drugs through methylation-independent transcription changes	Suzuki et al. (2002)
Methylation-sensitive restriction enzyme –sequencing	Methylation-sensitive digestion	Methylation-sensitive digestion, followed by adaptor ligation to cut-sites, sequencing	Unmethylated regions	Second-generation sequencing	Identifies unmethylated CpGs in recognitions sites of methylation-sensitive restriction enzymes, high coverage, increased genome-wide performance combined with MeDIP-seq	Maunakea et al. (2010) and Li et al. (2015)
*Hpa*II tiny fragment enrichment by ligation-mediated PCR (*Hpa*II tiny fragment enrichment by ligation-mediated PCR)	Methylation-sensitive digestion and isoschizomer (*HpaI*/*Msp*I)	Digestion, ligation mediated amplification, cohybridization, or sequencing	Differential display of unmethylated and methylated regions	Microarrays, second-generation sequencing	Simple, well-suited for identification of hypomethylated regions, less sensitive to genetic polymorphisms, requires enzyme recognition sites	Khulan et al. (2006)
Methylated DNA (MeD) immunoprecipitation	5-Methylcytosine antibody	Sonication, library preparation, immunoprecipitation, amplification, sequencing	Methylated regions	Oligonucleotide tiling arrays, second-generation sequencing	Genome-wide analysis, not dependent on presence of restriction enzymes, insensitive to genetic polymorphisms, bias toward CpG-rich regions (computational correction required), sensitivity is still discussed	Weber et al. (2005), Down et al. (2008), and Taiwo et al. (2012)
Methylcytosine DNA-binding domain (MBD) proteins	DNA-binding domain (MBD2) protein affinity selection	Sonication, immunopurification, library preparation, sequencing	Methylated regions	CpG island microarray, second-generation sequencing	Sequence specificity of MBDs, Potentially applicable to second-generation sequencing	Serre et al. (2010)

Methylated CpG island recovery assay	DNA-binding domain (MBD2/MDB3L1) protein affinity selection	Sonication, immunopurification, library preparation, sequencing	Methylated regions	Oligonucleotide tiling arrays, second-generation sequencing	Genome-wide analysis, not dependent on presence of restriction enzymes, insensitive to genetic polymorphisms, might have higher sensitivity than MeDIP	Rauch and Pfeifer (2005) and Jung et al. (2015)
Illumina HM450 array	Infinium assay	Bisulfite conversion, hybridization, single-base primer extension	Targeted differentially methylated CpGs	Microarrays	Genome-wide analysis of 485,000 selected CpG positions	Bibikova et al. (2011) and Sandoval et al. (2011)
Illumina HM EPIC array	Infinium assay	Bisulfite conversion, hybridization, single-base primer extension	Targeted differentially methylated CpGs	Microarrays	Genome-wide analysis of 840,000 selected CpG positions	Moran et al. (2015)
Whole-genome bisulfite sequencing, MethylC-seq	C/T base differences in second-generation sequencing	second-generation sequencing library preparation, bisulfite conversion, sequencing	Genome-wide, unbiased	Second-generation sequencing	Genome-wide analysis, CpG and non-CpG cytosines, genetic polymorphisms can be analyzed simultaneously	Kulis et al. (2012) and Urich et al. (2015)
Whole-genome bisulfite sequencing, postadaptor bisulfite tagging	C/T base differences in second-generation sequencing	Bisulfite conversion, second-generation sequencing library preparation, sequencing	Genome-wide, unbiased	Second-generation sequencing	Genome-wide analysis, CpG and non-CpG cytosines, genetic polymorphisms can be analyzed simultaneously	Miura et al. (2012)
Tagmentation-based whole-genome bisulfite sequencing	C/T base differences in second-generation sequencing	Transposase-mediated adaptor insertion, bisulfite conversion, sequencing	Genome-wide, unbiased	Second-generation sequencing	Genome-wide analysis, CpG and non-CpG cytosines, genetic polymorphisms can be analyzed simultaneously	Adey and Shendure (2012)
Reduced representation bisulfite sequencing	C/T base differences in second-generation sequencing	Restriction digest, second-generation sequencing library preparation, bisulfite conversion, sequencing	CpG islands, unbiased	Second-generation sequencing	Genome-wide analysis, CpG and non-CpG cytosines, focused on CpG-rich regions, cost-effective, genetic polymorphisms can be analyzed simultaneously	Meissner et al. (2005)

MethylC-seq; *PCR*, polymerase chain reaction.

(Mund et al., 2005; Kantarjian et al., 2006; Liu et al., 2007) to investigate the efficacy of novel demethylating agents (Balch et al., 2005), detect and predict the outcome of various human cancers (Hur et al., 2014; Inamura et al., 2014), and detect and quantify the effect of environmental exposure on the DNA methylome (Bollati et al., 2007; Marques-Rocha et al., 2016).

In situ hybridization methods with antibodies directed against 5-methylcytosine allow measurement of the methylation content and its potentially cell type−specific distribution, and thus to visualize 5-methylcytosine and its oxidative derivatives in their cell nuclear context (Miller et al., 1974; Bollati et al., 2007; Marques-Rocha et al., 2016). Because only clustered methylated CpGs prevalent in, for example, repeat elements can be recognized at the chromosomal level, methylation patterns at relatively small loci such as CpG islands contribute little to the overall staining profile. Therefore, technologies using an ultrasensitive and rapid fluorescence scanning system with submicrometer resolution have been devised to achieve the detection of methyl groups at specific promoters isolated from genomic DNA by restriction digestion and hybridization to capture oligonucleotides immobilized on a glass slide (Pröll et al., 2006).

7.3.2 Genome-Wide Analysis of DNA Methylation Patterns

Microarray-based technologies using small oligonucleotides with a length between 20 and 80 base pairs (bp) have been valuable methods for the identification of new epigenetic biomarkers because they do not require a priori knowledge of target DNA sequences. Their power to identify methylation markers for (early) diagnosis and classification of tumors has repeatedly been proven (Adorjan et al., 2002; Cottrell and Laird, 2003; Laird, 2003; Ushijima, 2005) and microarray-derived DNA methylation-based signatures were useful for predicting disease progression and risk of relapse (Wei et al., 2006) as well as response to specific cytotoxic drugs (Maier et al., 2005). With the exception of the epigenotyping arrays detailed subsequently, these technologies now have only historic value; they have contributed significantly to our current knowledge on DNA methylation but have been largely superseded by massively parallel sequencing, generally termed next-generation sequencing (NGS) (see Chapter 9), which, in contrast to microarray-based analyses, allows for a truly genome-wide analysis.

7.3.2.1 Methylation Reversal Approach

A method that sets itself slightly apart is the methylation reversal approach, which identifies methylation patterns not by analyzing changes in DNA methylation patterns themselves but by measuring mRNA expression after treating cells with a DNA methylation inhibitor (Suzuki et al., 2002). Pharmacologic agents such as the nucleoside analog 5-aza-2′-deoxycytidine (5-aza-dC) inhibit DNA methylation by forming a covalent complex between methyltransferase and DNA, which leads to progressive global demethylation (Yoo and Jones, 2006). This approach is therefore restricted to cell lines that can be cultured. Cells can be treated with 5-aza-dC alone or combined with a histone deacetylase inhibitor such as trichostatin A that leads synergistically to gene activation of epigenetically silenced genes (Sun et al., 2011). This approach uses well-established methods to analyze gene expression (microarrays or NGS) as a detection platform. In side-by-side comparisons, a sequencing-based readout proved to be more sensitive than microarray and correlated better with the quantitative polymerase chain reaction (qPCR) results for selected genes (Xu et al., 2013). The main advantage of this approach is that the detection of hypermethylated sites is linked to their transcriptional status, which allows identification of genes whose transcription is putatively under methylation control. However, modified expression patterns owing to changes of methylation in *cis* such as demethylation of the promoter, enhancer, or other regulatory elements of a target gene or changes owing to the activation of a gene that is in the same pathway as the target gene will be indistinguishable from each other. A further complication is the multiple effects of drugs such as 5-aza-dC on additional numerous metabolic pathways (Stresemann et al., 2006), which might lead to methylation-independent changes in gene expression.

7.3.2.2 Methylation-Sensitive and -Dependent Restriction Enzyme-Based Methods

Methylation-sensitive restriction endonucleases, that is, enzymes that are blocked by methylated cytosines in their recognition sequence (Bird and Southern, 1978), have been widely used to analyze methylation patterns in combination with their methylation-insensitive isoschizomers. About 50 MS restriction enzymes are known but few are available in combination with a methylation-insensitive isochizomer. One of the most commonly used pairs of enzymes is *Hpa*II/*Msp*I; both recognize and cleave the four-base palindrome C|CGG in double-stranded DNA, but whereas *Msp*I cleaves the DNA-independent of its methylation status, *Hpa*II is unable to cleave when the second cytosine is methylated (C|^{me}CGG). Another frequently used enzymatic combination is the MS enzyme *Sma*I (CC^{me}C|GGG), which leaves blunt ends, and *Xma*I (C|C^{me}CGGG), which is less sensitive to methylation and creates an overhang that can be used for the ligation of adaptors for subsequent PCR

amplification. Although methods based on MS restriction enzymes are relatively cost-effective because they require no special instrumentation, they are hampered by limitation to specific restriction sites, because only CpG sites found within these sequences can be analyzed. For example, only about 4% of CpG sites in nonrepetitive sequences are in recognition sites for the frequently used restriction enzyme *Hpa*II and only 0.03% can be cleaved by *Not*I (Fazzari and Greally, 2004). In addition, methods using these enzymes might be prone to false-positive results owing to incomplete cleavage, and some sequences are intrinsically resistant to digestion if not appropriately controlled. For example, non-CpG methylation on cytosines or DNA adducts in the vicinity of the cleavage site might influence the restriction capacity of an enzyme. Therefore, digestions are difficult to perform on material extracted from formaldehyde-fixed, paraffin-embedded (FFPE) material. Thus, approaches such as MS arbitrarily primed PCR and MS restriction fingerprinting (Gonzalgo et al., 1997, Huang et al., 1997, differential methylation hybridization (DMH) (Huang et al., 1999), MS representational difference analysis (Ushijima et al., 1997), methylated CpG island amplification (Toyota et al., 1999), microarray-based integrated analysis of methylation by isoschizomers (Hatada et al., 2006), and restriction landmark genomic scanning (Costello et al., 2002) have largely become obsolete and been replaced by more quantitative sequencing methods allowing rapid identification of altered DNA methylation levels at higher spatial resolution.

However, a few protocols are still in use and some restriction enzyme–based strategies have been successfully transferred to NGS instruments, which will permit a more quantitative analysis of the isolated (methylated or unmethylated) fraction of the genome. In addition, NGS does not restrict the analysis to a subset of loci of the genome as do microarrays, and the sequence identity is resolved in contrast to gel-based readouts.

Methylation-sensitive restriction enzyme sequencing (MRE-seq) identifies unmethylated CpG sites at single–CpG site resolution by sequencing size-selected fragments from parallel DNA digestions with a number of MS restriction enzymes (e.g., *Hpa*II, *Hin6*I, and *Aci*I) and covers about 1.7 million of the 29 million CpG sites in the human genome (Maunakea et al., 2010). After restriction digest and size selection, Illumina adaptors are ligated to the DNA fragments, PCR amplified, and sequenced using short-length (50-bp) single-read sequencing, thereby identifying unmethylated CpG sites within the restriction sites with single-base resolution. Interrogation of the methylated fraction of the genome by methylated DNA immunoprecipitation sequencing (MeDIP-seq), described in more detail subsequently, yields complementary information, and the combined use of MRE-seq and MeDIP-seq allows for genome-wide DNA methylation analysis at high coverage and resolution while limiting the biases of each technology (Li et al., 2015). A protocol similar to MRE-seq, methylation mapping analysis by paired-end sequencing, isolates both the methylated and unmethylated fraction of the genome by using the methylation-dependent restriction endonuclease McrBC and a combination of several restriction enzymes, respectively before library preparation and paired-end sequencing (Edwards et al., 2010).

In the initial protocol of differential methylation hybridization (Huang et al., 1999), two differentially treated fractions of the same sample were comparatively hybridized to an array-based probe library and relative (fluorescent) intensities were measured. Genomic DNA was fragmented by digestion with a frequently cutting restriction enzyme that preferentially cuts outside CpG islands. Linkers for PCR amplification were subsequently ligated to the digestion products. The sample was split into two parts and one half was digested with an MS restriction enzyme. In this sample only methylated fragments that are resistant to the digestion are amplified in the subsequent PCR, whereas in the reference sample all fragments are amplified. Fractions are labeled with two different fluorescent dyes and hybridized to microarrays with an immobilized CpG island library. The *Hpa*II tiny fragment enrichment by ligation-mediated PCR (Khulan et al., 2006) uses the digestion of this sample with its methylation-insensitive isoschizomer instead of a reference sample, which normalizes the signal for the *Hpa*II digest by that of the *Msp*I digest. This assay format has been successfully transferred to sequencing platforms and has been used in a number of studies (Hu et al., 2014; Yuan et al., 2016). Of note, a variation in the protocol permits the separate detection of 5-hydroxymethyl cytosine (Bhattacharyya et al., 2013).

Although for the analysis of gene-specific methylation patterns or individual CpG positions methods using MS endonucleases have largely been replaced by PCR-based methods following treatment of genomic DNA with sodium bisulfite, some are as described later still used and do provide some advantages when combined with high-throughput qPCR and multiplex amplification systems.

7.3.2.3 Methylated DNA Immunoprecipitation and Methyl Binding Protein Affinity Chromatography

The approaches described previously rely on MS restriction endonucleases and analysis is restricted to sequences that contain recognition sequences for the enzymes in their proximity. An alternative approach isolates the methylated fraction of a genome by immunoprecipitation or affinity purification of methylated DNA with MBD proteins. Of the different MBD proteins, MBD2b has the highest affinity for methylated DNA (Fraga et al., 2003). Therefore, in the methylated-CpG island recovery assay (MIRA), a

glutathione *S*-transferase–tagged full-length MBD2b has been used to bind sonicated methylated DNA fragments and the affinity to methylated CpG dinucleotides is further enhanced in a dose-dependent manner by the addition of the MBD3-like-1 protein (Rauch and Pfeifer, 2005). The combined effect significantly improves the sensitivity of the assay and a single methylated CpG dinucleotide enabled capture of the corresponding DNA molecule. Ligation of oligonucleotide linkers to enzymatically digested DNA before affinity chromatography permits efficient amplification of eluted fractions and subsequent analysis of differentially labeled input DNA and MIRA-enriched amplification products by DMH on a CpG island microarray (Rauch et al., 2006) or second-generation sequencing (Jung et al., 2015). Of note, because the capture is performed on double-stranded DNA, library preparation can be performed both before and after the enrichment step. A potential complication is the target specificity of MBDs, and MBD columns enrich significantly but do not fully purify methylated sequences (Selker et al., 2003). In the comparable MBD-seq or MethylCap-seq assay, only MBD2 is used to enrich methylated CpGs (Brinkman et al., 2010; Serre et al., 2010).

An alternative approach, MeDIP, follows the protocol for chromatin immunoprecipitation analysis, enriching methylated sequences independent of their surrounding sequence. Methylated DNA fragments are precipitated with a bead-immobilized antibody specific for 5-methylcytosine and analyzes locus specifically by PCR amplification (Luo and Preuss, 2003), genome-wide on various microarrays (Weber et al., 2005), or by sequencing (Down et al., 2008; Taiwo et al., 2012). The combination of MeDIP with tiling arrays was used to define the methylome of human promoters (Shen et al., 2007; Weber et al., 2007) as well as 13 human somatic tissues identifying tissue-specific methylated genes (Rakyan et al., 2008). MeDIP enriches preferential sequences with a high density of methylated CpGs such as repetitive elements or methylated CpG islands (Irizarry et al., 2008). Computational algorithms are therefore required to account for this bias in amplification and to convert signal intensities into the percentage of methylation (Down et al., 2008; Pelizzola et al., 2008; Huang et al., 2012; Xiao et al., 2015).

MeDIP-seq combines methylation-specific immunoprecipitation with sequencing of the pull-down product (Down et al., 2008). Because the antibody recognizes methylcytosine only in single-stranded DNA, sequencing adaptors are normally ligated to the fragmented DNA before immunoprecipitation and the isolated DNA fragments are subsequently amplified by PCR before sequencing. The resulting short reads are then mapped back onto the genome sequence. Stacked sequences indicate regions of the genome where the cytosines are methylated. Improved and automated protocols allow for the standardized high-throughput analysis of samples with little starting material (Taiwo et al., 2012).

Both MBD-seq and MeDIP-seq are well-suited for the identification of differentially methylated regions but have a different target distribution (Nair et al., 2011). Whereas MBD-seq preferentially enriches CpG islands and regions with high CpG density, MeDIP also enriches methylated CpGs in regions with low CpG density (Robinson et al., 2010; Nair et al., 2011). Of note, the CpG density eluted for sequencing using MBD-seq approaches can be modulated by using different salt concentrations for elution (Serre et al., 2010).

7.3.2.4 Genome-Wide Bisulfite Sequencing

Sodium bisulfite conversion of genomic DNA has revolutionized the field of DNA methylation analysis because it permits the use of well-established DNA amplification procedures from minute amounts of samples such as clinical specimens (Fig. 7.4) (Frommer et al., 1992). Bisulfite treatment of genomic DNA samples results in the hydrolytic deamination of nonmethylated cytosines to uracils, whereas methylated cytosines are resistant to conversion (Shapiro et al., 1974; Wang et al., 1980). After PCR amplification, in which uracils are replaced by their DNA analogue thymine, the methylation status at a given position is manifested in the ratio C (former methylated cytosine) to T (former nonmethylated cytosine) and can be analyzed as a virtual C/T polymorphism in the bisulfite-treated DNA. This chemical conversion of DNA is required to "freeze" the methylation status of a sample because DNA polymerases used for PCR amplification do not distinguish between methylated and nonmethylated cytosines, and therefore the methylation information is not retained after amplification. Sodium bisulfite treatment therefore enables the analysis of any CpG site of interest with quantitative resolution at the nucleotide level displaying CpG methylation as a positive signal.

Although potentially confounded by the presence of 5-hmC, whole-genome bisulfite sequencing (WGBS) or MethylC-seq can be considered the current reference standard for genome-wide identification of differentially methylated regions at single-nucleotide resolution. It overcomes the limitations of cloning and Sanger sequencing, a low-throughput method limited to a limited number of loci of interest, in which the quantitative resolution was limited by the number of clones analyzed (in most studies less than 20). Furthermore, WGBS avoids the problem with primer design that often introduces multiple biases (Grunau et al., 2001; Warnecke et al., 2002). The unprecedented quantitative and spatial resolution that is currently transforming DNA methylation analysis comes at a high cost and requires substantial sequencing to obtain proper and even coverage and requires some bioinformatic expertise and resources. The most widely used

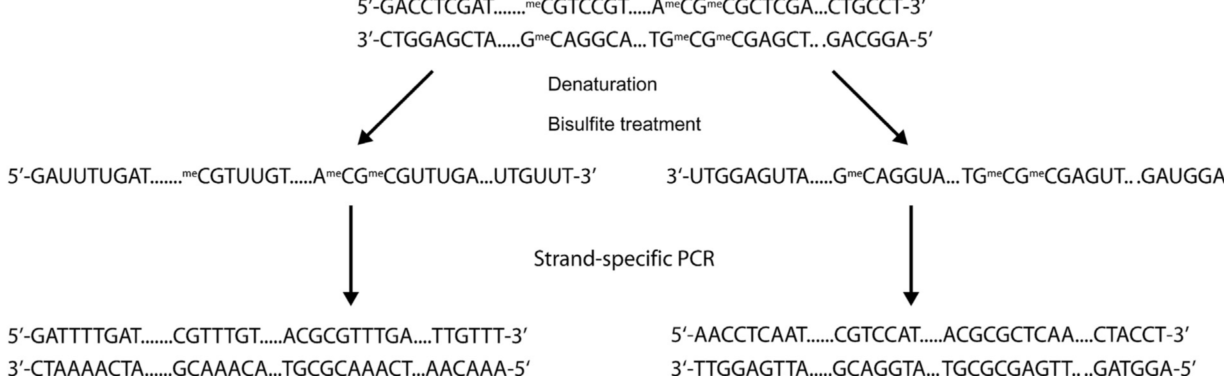


FIGURE 7.4 Principle of the treatment of genomic DNA with sodium bisulfite. Unmethylated cytosines are hydrolytically deaminated to uracils, which in turn are replaced by the DNA analog thymine during subsequent polymerase chain reaction (PCR) amplification. Methylated cytosines are resistant to the treatment under the carefully controlled reaction conditions and remain cytosines. Therefore, any cytosine still present in the bisulfite-treated DNA corresponds to a methylated cytosine before conversion. The Watson and Crick strands that are complementary in native genomic DNA are no longer complementary after bisulfite treatment and both strands can be independently amplified by PCR. This process doubles the genome size while reducing its complexity to a three-letter code, which makes primer design more challenging on bisulfite-treated DNA than on untreated genomic DNA.

protocol consists of the fragmentation of genomic DNA, adapter ligation, bisulfite conversion, and limited amplification using adapter-specific PCR primers. Whereas initially several micrograms of DNA were required to perform WGBS, the replacement of electrophoretic steps and gel extraction by magnetic beads has enabled the creation of libraries suitable for sequencing from about 100 ng of input material (Urich et al., 2015). Libraries have been reported to be constructed from less input material, but in most cases they require a high number of PCR cycles (up to 25 cycles) (Kobayashi et al., 2012), which potentially introduces a large bias for estimation of the DNA methylation levels and a much more substantial sequencing effort to obtain homogeneous and sufficient coverage. To assess bisulfite conversion efficiency, DNA bacteriophage λ is spiked into the reaction. Mapping the reads against the bisulfite-converted genome of the phage and counting any remaining cytosines allows problems to be identified during bisulfite conversion and estimate conversion rates. Similarly, overconversion can be assessed using a fully methylated spike-in. A large number of programs have been developed to perform quality control, preprocessing steps (such as adaptor, barcode, and quality score trimming), mapping of the reads to a bisulfite-converted reference genome and scoring of DNA methylation levels (count statistics), and identification of differentially methylated CpG positions and regions (Adusumalli et al., 2015). The protocol has been widely used for the methylome-wide analysis of a large number of organisms and plats as well as human tissues; about 90–95% of cytosines present in the genome are covered (Lister et al., 2008, 2009, 2011; Li et al., 2010; Lyko et al., 2010; Chalhoub et al., 2014; Guo et al., 2014).

Although most commonly performed on the Illumina sequencing platform, which allows higher coverage, protocols and analytical pipelines have also been devised for the SOLiD platform (Bormann Chung et al., 2010; Kreck et al., 2012) and have been applied in some studies (Hansen et al., 2011; Kreck et al., 2013).

A variation of the MethylC-seq protocol, which captures all four strands of a bisulfite-treated genomic DNA using an alternative sequencing adaptor strategy, has been devised (BS-seq) (Cokus et al., 2008). Although it allows a more comprehensive mapping of cytosines, the strategy requires a large amount of input DNA and comes at the cost of a more complex bioinformatics analysis; the approach has rarely been used (Popp et al., 2010). Similarly, whole-genome preamplification of bisulfite-treated DNA has been proposed as an alternative to obtain sufficient material for sequencing (Kobayashi et al., 2012); however, the accuracy and reproducibility of whole-genome amplification are still under debate and prone to bias especially when low amounts of input DNA are used (Bundo et al., 2012).

Tagmentation is based on a hyperactive variant of the prokaryotic Tn5 transpose that randomly fragments DNA and tags ends with the sequencing adaptors that can be used subsequently for amplification. Initially used for low-input genome sequencing (Adey et al., 2010), the approach has been adapted to the analysis of genome-wide DNA methylation analysis (Tn5mC-seq or T-WGBS) (Adey and Shendure, 2012; Wang et al., 2013), single-cell RNA sequencing (Brouilette et al., 2012), and as described subsequently, the analysis of chromatin accessibility (Buenrostro et al., 2013) and transcription factor binding sites and histone modifications (Schmidl et al., 2015). Because tagmentation requires double-stranded DNA as the target of the transposition reaction, tagmentation is performed before bisulfite conversion. Transposase complexes are loaded with methylated oligonucleotide (except for the 19-bp transposase recognition sequence) to retain sequence identity after bisulfite treatment and enable use of the standard amplification primers. A second complementary methylated adaptor is

added, replacing the transposase recognition sequence, and ligated using gap repair. Double-stranded DNA fragments are bisulfite converted and subsequently PCR amplified to append the flow cell–compatible primers. The protocol can be carried out with little input (down to 10 ng) because the tagmentation step removes the need for multiple steps of the conventional protocol for library preparation (DNA shearing, 3′-end repair, adenylation, and adapter ligation). Furthermore, transposome tagmentation has been shown to be little affected by guanine-cytosine (GC) content (Adey and Shendure, 2012). Up to 96% of CpGs can be covered with this approach and about 70% of reads align to the genome, a number slightly lower compared with the standard MethylC-Seq protocol. However, the coverage has been shown to be slightly more uniform compared with standard MethylC-seq and the tagmentation-based protocol requires up to about 100 less DNA input for similar sequence quality. Despite these differences as well as the more streamlined procedure, MethylC-seq and T-WGBS yield highly similar results in terms of methylation levels and coverage of the genome, and there seems to be no sequence bias for the insertion of the transposase (Wang et al., 2013). Furthermore, the presence of unmethylated nucleotides during the gap repair step serves as an internal control for bisulfite conversion efficiency and abolishes the need for DNA spike-ins. Of note, T-WGBS has been found to be unsuitable for the analysis of DNA extracted from FFPE tissue.

A potential drawback of both the MethylC-seq and the tagmentation-based protocol is that the adaptors are ligated to the DNA fragments before the bisulfite conversion. However, treatment of DNA with sodium bisulfite leads to a substantial degradation in DNA and significantly reduces the amount of amplifiable DNA through the induction of double-strand breaks between the adaptors. Therefore, protocols performing the adaptor tagging after bisulfite treatment [postadaptor bisulfite tagging (PBAT)] have been devised and shown to enable efficient library construction from as little as 125 pg of DNA (Miura et al., 2012). This approach has therefore been used to sequence methylomes of rare cell population, such as primordial germ cells or zygotes (Kobayashi et al., 2012, 2013; Shirane et al., 2013; Peat et al., 2014). Adaptor tagging is performed with two rounds of random primer extension using oligonucleotides with a random tetramer sequences at the 3′-end of amplification primers containing the Illumina adaptor sequences. Primers are biotinylated for the capture of biotinylated fragments after first-strand synthesis on magnetic beads. Starting from about 100 ng of DNA allows routine PCR-free construction of libraries for methylome-wide sequencing (Miura et al., 2012), thereby avoiding the problem of high PCR duplicate rates that frequently occurs in PBAT protocols owing to the preferential binding of the random amplification primers when the library has insufficient complexity and diversity. Furthermore, random primers tend to amplify sequences preferentially with elevated GC content. Whereas the classic MethylC-seq tends to cover GC-poor regions better than GC-rich regions, this phenomenon seems to be reversed in the PBAT protocol, which suggests that if PCR amplification is required, the combination of the two approaches will probably yield the most even coverage of the methylome.

With the drop in the cost of NGS, large-scale projects such as the International Human Epigenome Project (ihec-epigenomes.net) and the National Institutes of Health Roadmap Epigenomics Project (www.roadmapepigenomics.org) have moved to applying WGBS to establish reference epigenomes of all cell types in the human body (Roadmap Epigenomics et al., 2015). This technique provides complete resolution; however, it is still too expensive to apply to the study of many individuals. A strategy to apply BS-seq to genome-wide studies in more complex organisms and across populations such as mice and humans consists of reducing the complexity of the genome by creating reduced representation libraries [reduced representation bisulfite sequencing (RRBS)] (Meissner et al., 2005). The methylation-insensitive restriction endonuclease *Msp*I followed by size selection is used to isolate small fragments originating from CpG-rich regions, significantly enriching for CpGs in CpG islands. This fraction of the genome is then comprehensively sequenced. However, the isolated fraction covers only 1% of the genome. Predesigned products enabling the capture of promoters and CpG islands are currently commercially available, such as Agilent's SureSelect™ human Methyl-seq, covering approximately 3% of the genome (84 Mb) corresponding to about 1/7 of all sequencing-accessible CpG sites (3.7 M CpG sites) (Borno et al., 2012). The technology also allows customized capture panels targeting regions of interest (Ivanov et al., 2013). Although the required amount of starting material has been significantly reduced since the method was originally devised [20–30 μg (Lee et al., 2011) to 2–3 μg], it remains a major drawback of these methods. This requirement is mainly because the capture is carried out after adapter ligation and before the bisulfite conversion proscribing an amplification step as well as the high number of PCR cycles that need to be performed after the release of the captured fragments potentially distorting the DNA methylation values. However, advances show the feasibility of a PBAT protocol significantly reducing the amount of starting material required (Miura and Ito, 2015). The design of custom capture arrays targeting a high number of specific regions of interest makes this approach competitive with RRBS, which costs substantially less but might be less well-suited if specific regions are mandatory for the analysis that are not easily accessible by a (combination of) restriction digests. The

SeqCap Epi CpGiant from Roche/Nimblegen allows for the reduced input of DNA (1 μg) compared with the SureSelect MethylSeq system and is performed on a WGBS library by using capture probes complementary to methylated, unmethylated, and partially methylated targets after bisulfite conversion. A sophisticated design combined with the use of long probes allows efficient capture focusing only on regions of interest at much increased coverage compared with WGBS (Allum et al., 2015). Up to four samples can be sequenced on a single lane of a flow cell of an Illumina HiSeq 2000, which is 30% less sequencing depth compared with the MethylSeq capture (70−80 million reads compared with 100−150 million reads). Read numbers as low as 40−50 million have been reported to be sufficient for coverage allowing accurate quantitative assessment.

7.3.2.5 Epigenotyping Arrays

Illumina, Inc. (http://www.illumina.com) presented a modified version of the Golden Gate genotyping assay for the analysis of DNA methylation. The Golden Gate assay analyzes 1505 CpGs in 807 genes using 250 ng of bisulfite-converted DNA as the input (Bibikova et al., 2006b). Allele discrimination is achieved before amplification: two allele-specific oligonucleotides that have their 3′ terminus complementary to the methylated or unmethylated alleles of a target CpG and that are extended only if the primer matches the target perfectly, and two locus-specific oligonucleotides that are also complementary to an unmethylated or methylated allele (Bibikova et al., 2006b). Extended allele-specific oligonucleotides are then ligated to the locus-specific oligonucleotides to create templates for subsequent amplification with fluorescently labeled primers specific for the respective methylation state. Allele discrimination is thus achieved at two different levels, which increases the specificity of the reaction and minimizes false-positive results. The locus-specific oligonucleotide also carries a sequence that is complementary to tag sequences immobilized on the bead arrays and that is then used for readout of the quantitative methylation degree at selected CpG positions with a resolution of about 20%. Results are normally confirmed by high-resolution quantitative methods such as pyrosequencing (Ladd-Acosta et al., 2007). The Golden Gate assay has been used to study the DNA methylation profile of embryonic stem cells (Bibikova et al., 2006a) and brain tissue (Ladd-Acosta et al., 2007), as well as to identify DNA methylation changes specific to Wilms' tumor (Bjornsson et al., 2007) and neuroendocrine tumors (How-Kit et al., 2015) and has until recently been continued as custom panel for multiplexed DNA methylation assays under the name of the VeraCode DNA methylation assay. However, the assay is based on the hypothesis that closely neighbored CpG positions display a high degree of co-methylation, which might be a simplified assumption and the coverage was of course very limited. Golden Gate has therefore seen three subsequent generations of epigenotyping arrays with ever increasing coverage. The Infinium assay analyzes for 27,578 CpG sites in more than 14,000 genes (Bibikova et al., 2009). Twelve samples are processed in parallel. Similar to the respective genotyping assay, the 3′ base is specific for the methylation state, which is extended only in the case of complete complementary with one of the four fluorescently labeled dideoxynucleotides. A different bead "barcode" distinguishes between probes specific for the methylated and unmethylated state of a given CpG. This assay allows a comprehensive analysis of the DNA methylation state of a sample in high-throughput studies, but again only one to three CpGs are analyzed per gene. With the Illumina HM450 array 485,000 CpG positions across the human genome can be interrogated using the Infinium assay. This array has turned into an industry standard for DNA methylation studies in humans (Bibikova et al., 2011; Sandoval et al., 2011) and has been rapidly adapted by the community for epigenome-wide association studies for the analysis of a large variety of diseases and phenotypes (Grundberg et al., 2013; Dick et al., 2014; Glossop et al., 2014; Martino et al., 2014; Miceli-Richard et al., 2015; Stefansson et al., 2015). The advantage of this method is the integration of a methylation-dependent step into a highly standardized protocol that has been thoroughly validated, and random variations in signal intensities owing to differential hybridization efficiencies have been minimized. When properly handled these epigenotyping arrays are accurate and display high correlation with RRBS (Bock et al., 2010) as well as locus-specific quantitative assays such as pyrosequencing (Roessler et al., 2012) or MethyLight (Campan et al., 2011). The Illumina MethylationEPIC BeadChip has been presented and evaluated; it adds approximately 400,000 CpGs to the content of the Illumina HM450 array and focuses essentially on enhancer regions identified in the ENCODE and FANTOM5 projects, thereby significantly extending the information content of the BeadChip (Moran et al., 2015).

The increasing information content of these dedicated epigenotyping arrays has rendered obsolete the analysis of the methylated fraction of the genome enriched by MS digestion on commercial single-nucleotide polymorphism (SNP) genotyping microarrays, a method termed methylation SNP (Yuan et al., 2006; Kerkel et al., 2008). An advantage was the simultaneous genetic (DNA copy number/loss of heterozygosity) and epigenetic analysis of a genome with dense coverage, but this information can now also be obtained directly from the epigenotyping arrays (Feber et al., 2014).

Genome-wide discovery methods are well-suited for mapping large-scale methylation variation in the whole

genome, but for some of them they do not provide quantitative high-resolution information of individual CpG sites (e.g., MeDIP-seq, MBD-seq) and/or can be performed only in a limited number of individuals because of the cost (WGBS). Therefore, the large number of candidate genes or regions potentially involved in biological or medical conditions under investigation needs to be validated using more quantitative methods in a larger sample set (potentially hundreds of samples).

7.3.3 Analysis of Candidate Gene Regions

The following methods allow overall density of DNA methylation to be analyzed in an amplified target region after bisulfite treatment (Fig. 7.4) (Frommer et al., 1992).

Methods for the analysis of the global methylation levels of a target region do not yield precise information about the methylation status of individual CpGs, and the distribution of allelic methylation patterns is only more or less resolved depending on the technique. These methods are valuable screening tools that might help select informative samples for more detailed analysis using techniques with increased resolving power such as sequencing, mass spectrometry−based sequencing, and pyrosequencing. Analysis of DNA methylation levels in candidate regions can be analyzed semiquantitatively by techniques such as denaturing high-performance liquid chromatography (Deng et al., 2002) or MS single-strand conformation analysis (Burri and Chaubert, 1999), all of which use differential retention or migration behavior owing to the sequence difference between former methylated and unmethylated molecules induced by bisulfite treatment. However, heterogeneous methylation levels might result in patterns that are difficult to interpret. Melting curve analysis (Fig. 7.5) monitors the melting temperature of a PCR product amplified from bisulfite-treated DNA in real time using a thermocycler coupled to a fluorometer (Worm et al., 2001; Akey et al., 2002). The melting profiles of PCR products originating from methylated and unmethylated variants of the same template are significantly different because of their different GC contents. Therefore, the methylation status of an unknown sample can be determined by comparing the melting profile of the sample with calibration standards. A gradual increase in temperature leads to stepwise dissociation of the double strand in domains of the PCR product in the function of their GC content that differs between methylated and unmethylated molecules after bisulfite treatment. The technology was previously hampered by the toxicity of the intercalating agent SYBR Green I to DNA polymerases, which prohibited working at the required saturating concentrations. Advances in fluorescence detection technology, new algorithms for data calculation, and the use of novel dyes have allowed the development of this high-resolution melting analysis (Wojdacz and Dobrovic, 2007). This method is a promising advance because it is extremely simple and relatively inexpensive, and it allows the rapid scanning of a large number of genes for the presence of differential DNA methylation. Diagnostic applications for the detection of aberrant methylation profiles in imprinting disorders (White et al., 2007; Wojdacz et al., 2008; Alders et al., 2009) and cancer (Balic et al., 2009) have been proposed.

7.3.4 Sequencing Technologies

7.3.4.1 Sanger Sequencing of Bisulfite-Treated DNA

Direct fluorescent Sanger sequencing of PCR products amplified from bisulfite-treated DNA can be used for the detailed analysis of DNA methylation in long CpG stretches from small amounts of starting material (Clark et al., 1994). However, direct sequencing is technically demanding in regions with heterogeneous or low (<15%) methylation levels. Furthermore, quantitation of the readout is imprecise and difficult because of the low sequence diversity of bisulfite-converted DNA, which requires sophisticated computational algorithms such as ESME (Lewin et al., 2004). Therefore, in most cases PCR products are cloned and multiple colonies are subsequently sequenced. Cloning and sequencing of multiple colonies (Frommer et al., 1992) were previously regarded as the reference method for the analysis of DNA methylation because of the richness of information provided, and most novel methods were validated by comparison with results obtained by sequencing and cloning. It provides detailed information on the methylation status of CpG positions in relatively long sequence stretches of target sequence and has been widely used in the past. Before the advent of second-generation sequencing, cloning and subsequent Sanger sequencing was the only technology capable of addressing the problem of allele-specific and/or mosaic methylation patterns adequately. However, the quantitative resolution is limited by the number of clones analyzed. Because the procedure becomes expensive, time-consuming, and cumbersome when applied in high throughput, only a small number of alleles on the order of about 10 to 20 colonies are usually analyzed, from which it is difficult to infer statistically meaningful results. Despite its status as a reference method, this procedure is prone to a variety of biases (PCR and cloning procedure) that distort the quantitative result. A variety of critical parameters as well as potential sources of artifacts and their remediation have been investigated in detail to avoid these pitfalls (Warnecke et al., 1997; Grunau et al., 2001).

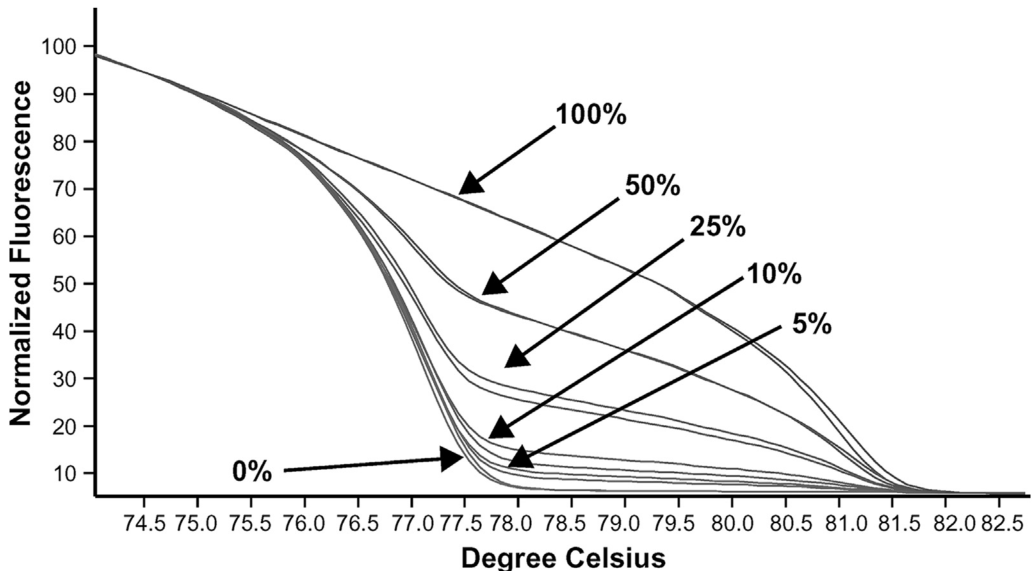


FIGURE 7.5 High-resolution melting analysis of different mixtures of methylated and unmethylated DNA (0%, 5%, 10%, 25%, 50%, and 100%) in the promoter region of *RARB* gene.

7.3.4.2 Next-Generation Sequencing of Bisulfite-Treated DNA

Much greater coverage and quantitative resolution can be obtained by NGS. Quantitative sequencing approaches make use of benchtop sequencers (e.g., Illumina's MiSeq; 454/Roche's Junior or Ion Torrent's PGM) and allow generating high levels of coverage (e.g., hundreds to thousands times) that will yield precise measurements of the quantitative levels of cytosine methylation. Amplicon bisulfite sequencing using these instruments has become widely used to validate genomic regions after methylome analyses and to answer hypothesis-driven research questions. In addition, because of the sequencing of clonal clusters generated in the sequencing machine, these methods provide co-methylation patterns on individual molecules within the limits of the length of the reads (up to 600 bp in paired-end mode). A current output of about 50 million reads for the MiSeq yields between 3.8 and 15 GB of sequence depending on the sequencing kit, and several tens to hundreds of target regions can be analyzed simultaneously depending on the desired coverage and number of samples analyzed in parallel. Their short run time, relatively low running costs, and wide availability make them a valuable alternative for targeted DNA methylation analysis. Although initially approaches were devised for the analysis on the pyrosequencing-based 454 platform (Taylor et al., 2007; Gries et al., 2013), which has been shown to yield accurate and quantitative results, and despite its nearly unrivalled sequencing length, this sequencing platform has not been able to cope with the improvements of other platforms and is supposed to be phased out in the near future.

In general, PCR amplification products are prepared from bisulfite-treated DNA using a two-round amplification protocol with a first pair of target region-specific amplification primers that contain tag sequences to label the created amplicons with sequences compatible for subsequent PCR amplification with full-length Illumina p5 and p7 adaptor sequences. Molecular barcodes and full-length adaptor sequences are added in a second round of amplification after pooling all amplicons from a sample. If sample quantity is limited, the first amplification can also be performed as multiplex PCR, but it requires some more optimization for the multiplex setup (Korbie et al., 2015). Another strategy consists of using conventional amplifications: primer and the molecular barcode and adaptors complementary to the sequences immobilized on the flow cell are added in a standard library preparation protocol after pooling of the PCR products from the first round of amplification (Jenkins et al., 2014). An alternative protocol, termed bisulfite amplicon sequencing, uses the hyperactive Tn5 transposase (Nextera/Illumina) for random insertion of the sequencing primers in the amplification product after PCR-based targeted amplification using conventional primers (Masser et al., 2013). Amplification of multiple amplification products in parallel can also be performed using microfluidic tools such as the Fluidigm Access Array simultaneously amplifying 48 target regions in 48 samples starting from as little as 50 ng (Paliwal et al., 2013). The Fluidigm Access Array uses a two-step amplification procedure in which universal forward and reverse adaptor sequences are added to the 5′-ends of the gene-specific amplification primers. Sample barcodes and platform-specific sequencing primers are added in a second round of amplification. However, this approach has been found to

be unsuitable for DNA extracted from FFPE samples (Korbie et al., 2015). Bisulfite-patch PCR is another approach enabling multiplex amplification by using a restriction enzyme digest to anneal exonuclease-resistant patch oligonucleotides and universal primers complementary to the created overhang, while unselected fragments are eliminated by an exonuclease digestion (Varley and Mitra, 2010). After bisulfite conversion, fragments are amplified using sequencing platform-specific universal primers. Bisulfite patch PCR has been used to analyze up to 94 simultaneously amplified sequences with little off-target sequences; however, the requirement for a specific restriction site imposes serious limitations on the targets that can be analyzed in parallel. Nonetheless, multiple restriction enzymes can be used to select different sets of target sequences (Varley and Mitra, 2010).

As for genome-wide approaches for DNA methylation analysis, PBAT can also be applied to amplicon bisulfite sequencing approaches to start with about 100 times less input DNA compared with the standard library preparation protocol and to reduce the number of amplification cycles (Miura and Ito, 2015). In general, bisulfite sequencing using next-generation sequencers with their digital readout enables more accurate quantification of DNA methylation levels because they show a reduced error of quantitation and lower standard deviations compared with conventional (analog) sequencing approaches (5% vs. 5−20%) (Masser et al., 2013). A sequencing depth of approximately 1000× is sufficient for precise measurement of the DNA methylation levels and increasing sequencing depth does not further improve accuracy. However, accuracy is already high (>99%) if only regions with a reasonable amount of coverage (i.e., >50×) are used for base calling and quantitative determination. Up to 96 samples analyzing multiple regions of interest can be analyzed in parallel using conventional multiplexing strategies such as (dual) indexing. The shift of several vendors from six base indexes to eight base indexes will further increase multiplexing capabilities. Of note, the MiSeq already exists in a version certified for diagnostics, which potentially allows the analysis of DNA methylation-based biomarkers in a clinical setting. Specialized bioinformatics pipelines provided with the instruments or freely available pipelines as Bismark (Krueger and Andrews, 2011) or BiQ-Analyzer (Lutsik et al., 2011) enable convenient and standardized analysis of the sequencing results including the demultiplexing of individuals, alignment to target regions, and estimation of the DNA methylation degree allowing the analysis of DNA methylation in target regions without the end user needing great bioinformatics expertise.

Ion Torrent's PGM sequencer has also been used for locus-specific DNA methylation analysis (Nones et al., 2014); this technology has been employed to sequence the methylation patterns of candidate genes in circulating cell-free DNA (Vaca-Paniagua et al., 2015).

7.3.4.3 Pyrosequencing

One method that has received much attention for the simultaneous analysis and quantification of the methylation degree of several CpG positions in close proximity is Pyrosequencing™ (Uhlmann et al., 2002; Colella et al., 2003; Tost et al., 2003a; Dupont et al., 2004; Tost and Gut, 2007). Pyrosequencing technology, in contrast to conventional Sanger sequencing, is based on sequencing by synthesis. A single nucleotide is added to a primer hybridized to a template strand and its incorporation is monitored by the luminometric detection of pyrophosphate that is released upon nucleotide incorporation and converted into a light signal by a cascade consisting of three enzymes (Ronaghi et al., 1998). Remaining nucleotides are degraded by the fourth enzyme, an apyrase, enabling the cyclic dispensation of nucleotides in a predefined order. One major strength of the technology is the quantitative nature of the results. The bioluminometric response is linear ($R^2 > 0.99$) for the sequential addition of up to five identical nucleotides (C, G, and T) or three deoxyadenosine 5′(α-thio)triphosphate. Pyrosequencing is therefore ideally suited for DNA methylation analysis after bisulfite treatment because it combines the ability of direct quantitative sequencing, reproducibility, speed, and ease of use (Tost and Gut, 2007). Information about an entire amplified region can be obtained by repeated stripping of the de novo synthesized strand from the template and hybridization of new sequencing primers (serial pyrosequencing) (Tost et al., 2006). Pyrosequencing provides quantitative information about the methylation status of all CpG positions in a sequence of interest and is therefore contrast with methylation-specific PCR (MSP) methods, which are also ideally suited for the analysis of regions displaying heterogeneous methylation levels. Another advantage of pyrosequencing compared with fluorescent (real-time) MSP methods is the detection of the sequence surrounding the polymorphic CpG positions instead of a fluorescent signal, which does not contain direct information about the actual amplified target. It also permits the detection of SNPs present in the bisulfite-treated sequence. Besides several studies on genes aberrantly silenced by promoter hypermethylation in cancer (Fig. 7.6), pyrosequencing has become the most widely used method and reference technology for quantitative DNA methylation analysis at single-nucleotide resolution to detect small changes in DNA methylation resulting from different environments, for example (Michel et al., 2013; Martinez et al., 2014).

7.3.4.4 Matrix-Assisted Laser Desorption/Ionization−Mass Spectrometry-Based Sequencing

An alternative method for sequencing PCR products amplified from bisulfite-treated DNA uses matrix-assisted laser desorption/ionization (MALDI)−time-of-flight mass

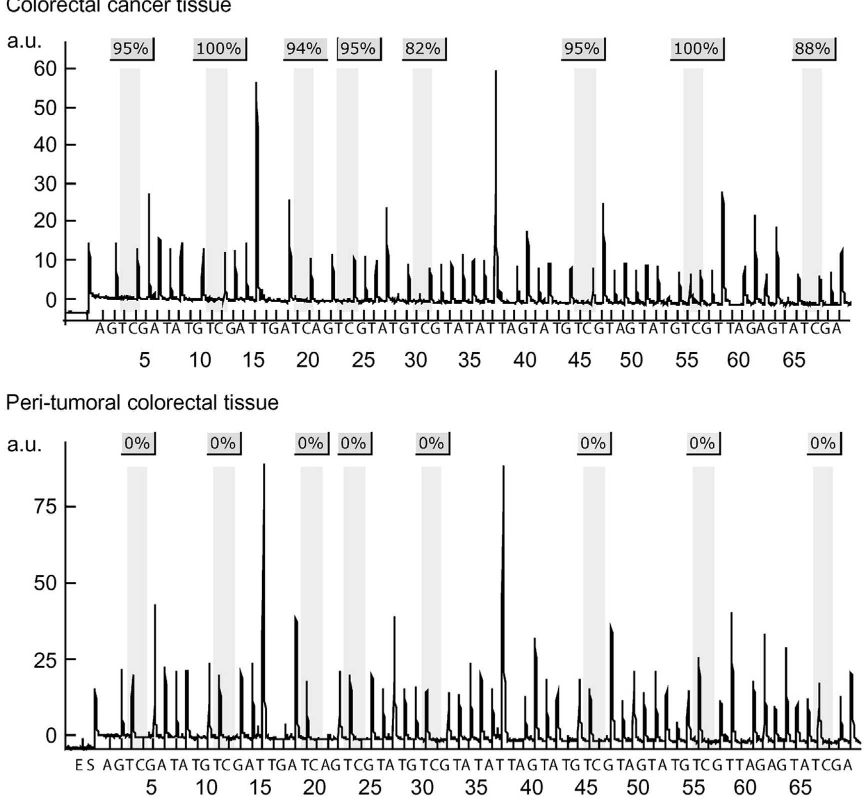


FIGURE 7.6 Pyrograms of the promoter region of the mismatch repair gene *MLH1*, which is commonly inactivated in microsatellite-instable colorectal cancer by promoter hypermethylation. Top: DNA extracted from the tumor. Bottom: DNA extracted from distant peritumoral tissue of the same patient.

spectrometry, for which fragile biomolecules are mixed with a matrix of small organic molecules absorbing most of the energy of the laser, which is used to volatilize the analyte and transfer part of it onto the analyte. MALDI has been proven useful for the complex analysis of nucleic acids such as human leukocyte antigen typing, haplotyping, gene expression, and DNA methylation analysis and tissue imaging (Tost and Gut, 2006). Because the direct analysis of DNA in a MALDI mass spectrometer is associated with some problems (Tost and Gut, 2002), bisulfite-treated DNA is transcribed into RNA, which is subsequently cleaved base-specifically using a variety of RNAses (Schatz et al., 2004; Ehrich et al., 2005). In Agena's (formerly Sequenom's) MassCLEAVE protocol, a modified T7 polymerase is used that does not discriminate between ribonucleotides and deoxyribonucleotides (Ehrich et al., 2005). Transcription is carried out with three ribonucleotides, and either cytidine triphosphate or uridine triphosphate is replaced by the respective noncleavable deoxynucleotide. RNAse A is used for the subsequent C- or U-specific cleavage. CpGs that are methylated differentially between two samples in the amplified region can be identified by shifts in the peak patterns resulting from the mass difference in a CpG dinucleotide containing cleavage product or the presence or absence of a peak corresponding to a certain fragment, respectively. The use of universal reverse transcription primers makes these procedures suitable for high-throughput applications with high quantitative resolution, as demonstrated on the analysis of more than 400 loci in 59 cancer cell lines (Ehrich et al., 2008). This approach can also be used to determine allele-specific DNA methylation patterns if SNPs are present within the analyzed fragment (Coolen et al., 2007). Potential complications might arise, however, if several fragments of the same mass are created by the cleavage; therefore, peaks cannot be assigned unambiguously. Compared with other techniques that are able to achieve quantitative DNA methylation data on consecutive CpGs in a region of interest, its quantitative resolution of about 5% and a similar limit of detection for the minor methylation allele fraction are rivaled only by pyrosequencing (Tost and Gut, 2007) and these described targeted bisulfite-sequencing approaches using NGS. Systematic evaluation of each step in the workflow showed that most of the variability of the experiment was induced by either the bisulfite treatment of the DNA or the subsequent PCR amplification, whereas the cleavage and the mass spectrometric analyses contributed much less to the variability in the observed quantitative measurements (Coolen et al., 2007; Ehrich et al., 2007). In more than 90% of the tested samples, a methylation difference of 10% was successfully detected with the mass spectrometric assay (Coolen et al., 2007). The procedure is also amenable to the

analysis of DNA extracted from fresh-frozen and FFPE tissues (Radpour et al., 2009).

The EpiTYPER is one of the most widely used methods for the analysis of gene-specific DNA methylation patterns and has been applied to the large-scale analysis of DNA methylation patterns in cancer (Radpour et al., 2009). Because of its high-throughput capacity measuring 96−384 PCR products in parallel, it is also one of the most widely used methods for the validation of DNA methylation variation with specific phenotypes identified in epigenome-wide association studies (Zeilinger et al., 2013; Tobi et al., 2014; Zhang et al., 2014).

7.3.5 Analysis of Individual CpG Positions

7.3.5.1 Restriction Enzyme-Based Analysis of Single CpG Positions

Southern blotting is a simple and inexpensive procedure for analyzing methylation status at specific CpGs by digesting genomic DNA with MS restriction enzymes and subsequent Southern analysis (Bird and Southern, 1978). This approach enabled some of the most intriguing findings at the beginning of DNA methylation research, such as the genome-wide hypomethylation of human cancers (Feinberg and Vogelstein, 1983). It provides authentic information about the methylation status of a CpG in an enzymatic recognition sequence because it does not rely on PCR amplification. However the trade-off is that a large amount of DNA (several micrograms) is required and it is not well-suited for a heterogeneous population of methylated molecules that might occur, for example, in cancerous tissues. Exponential amplification by PCR amplification after MS restriction digestion is an alternative that requires substantially less DNA and no prior bisulfite conversion treatment, which makes it well-suited as a rapid screening tool for differential methylation (Singer-Sam et al., 1990). Multiple targets can be analyzed simultaneously by locus-specific multiplex PCR after MS restriction digest of genomic DNA (Melnikov et al., 2005). Additional information on the methylation status of a target region can be achieved by digesting the DNA with either MS restriction enzymes or methylation-dependent enzymes such as *McrBC*, which distinguishes complete methylation, partial methylation, and the absence of methylation in the sequence (Yamada et al., 2004). Quantification can be improved by monitoring the increase in fluorescence by quantitative real-time PCR with the intercalating dye SYBR Green (Bastian et al., 2005; Oakes et al., 2006). Combined with microfluidic preparation of the PCR products, this method allows the analysis of a large number of target sequences from a limited amount of starting DNA (Wielscher et al., 2015). Complete digestion of methylated DNA is ensured by overdigestion of the DNA, a combination of several restriction enzymes, as well as the requirement of two or preferably three restriction sites within the target sequence to minimize false-positive results resulting from incomplete digestion.

A method that has attracted a lot of interest for diagnostic applications is MS multiplex ligation-dependent amplification (Nygren et al., 2005). Initially devised for the analysis of copy number alterations, this technique has been useful for the parallel analysis of up to 40 loci for comprehensive analysis of all possible variations of DNA methylation aberrations in imprinting disorders (Dikow et al., 2007; Priolo et al., 2008), the combined analysis of genetic and epigenetic alterations in imprinting disorders (Scott et al., 2008), and tumor analysis (Berkhout et al., 2007). Two oligonucleotides with universal primer binding sites are annealed to a target region and ligated in the case of complete target complementarity. An MS enzyme is added to the ligation reaction, digesting unmethylated templates and reducing the amount of ligated product. A semiquantitative readout is then performed using capillary electrophoresis.

Another method for the rapid screening of samples relies on the creation of new restriction enzyme recognition sites through bisulfite treatment. Bisulfite-converted DNA samples are PCR-amplified after restriction digestion and methylation is analyzed through the presence or absence of bands on an agarose gel (Sadri and Hornsby, 1996). The quantitative accuracy of the assay can be improved when analysis is carried out by polyacrylamide gel electrophoresis with subsequent electroblotting, hybridization with labeled oligonucleotides, and quantitation using a phosphoimager [combined bisulfite restriction analysis (COBRA)] (Xiong and Laird, 1997). Although simple and inexpensive, COBRA becomes labor intensive when quantitation is required. The latter point has been greatly improved by implementation of the Agilent Bioanalyzer as a detection platform (Brena et al., 2006). The Bioanalyzer separates and quantifies DNA fragments via electrophoresis in microfluidic chips for the direct quantitative visualization of restriction products created by COBRA.

7.3.5.2 Methylation-Sensitive Single-Nucleotide Primer Extension Assays

The methylation degree at given CpG positions is manifested by the ratio C (former methylated cytosine) to T (former nonmethylated cytosine) after bisulfite treatment and can be analyzed as a virtual C/T polymorphism after a strand-specific PCR. Therefore, quantitative and accurate information about individual methylation-variable positions can be obtained by MS single-nucleotide primer extension (MS-SNuPE). A PCR template is generated irrespective of the methylation status of the target region; internal primers are hybridized to the target sequence and terminate immediately 5′ of the CpG position to be assayed. Extension is carried out with (labeled) terminating dideoxynucleotides. Analysis of individual CpG positions requires

prior knowledge of which CpG positions are of interest because CpG islands can contain hundreds of CpGs, for example. However, once these positions are known, methods based on primer extension are well-suited to high-throughput analysis of a few CpGs in large sample cohorts, enabling epigenetic analyses on an epidemiologic scale. The close proximity of potentially polymorphic positions in CpG-rich regions such as CpG islands complicates the design of SNuPE assays irrespective of the detection platform used, because this might cause preferential annealing of the primers to a subpopulation of methylated molecules displaying a specific methylation pattern. A variety of detection platforms have been combined with MS-SNuPE (Table 7.2). Mass spectrometry—based assays (Tost et al., 2003a; Ragoussis et al., 2006) permit higher levels of multiplexing and were applied in a multiplexed form as a reference method for verification and quantitative fine typing of the results obtained by direct BS-seq for the pilot study Human Epigenome Project (Rakyan et al., 2004). However, they require more sophisticated instrumentation in contrast to capillary electrophoresis—based methods such as MS-SNaPshot (Uhlmann et al., 2002; Kaminsky et al., 2005).

7.3.6 Methylation-Specific Polymerase Chain Reaction and Real-Time Polymerase Chain Reaction Methods

DNA methylation has received a lot of attention because of its potential as a stable and amplifiable biomarker for early diagnosis, prognosis, or response to treatment in various types of cancer (How-Kit et al., 2012). Biomarkers capable of distinguishing cancerous cells from normal ones must be specific, sensitive, and detectable in specimens obtained through minimally invasive procedures to be clinically applicable. MSP (Herman et al., 1996) and methylation-specific real-time PCR-based methods such as MethyLight (Eads et al., 2000a), HeavyMethyl (Cottrell et al., 2004) or quantitative analysis of methylated alleles (QAMA) (Zeschnigk et al., 2004) have been proven to be very well-suited for detecting cancer-specific methylation patterns in primary tumor tissues and, more important, for detecting low levels of methylation in circulating DNA (Fig. 7.7). Tumor-derived methylated DNA molecules can be found in various body fluids such as urine, sputum, and serum or plasma of patients with cancer. Detection and monitoring of these methylated molecules in the presence of an excess of normal (and usually unmethylated) DNA can be achieved with various approaches, as described subsequently. Because the amplicons of the different MSP variants are small (approximately 100 bp), these methods usually work well with DNA of lower quality, such as DNA extracted from FFPE samples (Herman et al., 1996). The similarity of the approaches described in the following sections to real-time expression analysis facilitates implementation of the technology, execution of the experiments, and interpretation of the results for laboratories not yet familiar with DNA methylation analysis. The design of assays and optimization of amplification are probably the most important part to ensure specific amplification of the desired locus. Sensitivity and specificity vary largely among assays depending on primers (and probes in case of techniques such as MethyLight) and conditions.

7.3.6.1 *Methylation-Specific Polymerase Chain Reaction*

MSP allows the amplification of virtually any CpG sites after bisulfite treatment with three pairs of primers, complementary to the former methylated sequences, the former unmethylated sequences, or genomic, unconverted DNA, respectively (Fig. 7.7) (Herman et al., 1996). The latter serves as a control for complete bisulfite conversion. Primers need to hybridize to sequences with at least two methylation variable positions (CpGs) to obtain the necessary specificity for selective amplification. The presence or absence of an amplification product analyzed on a conventional agarose gel reveals the methylation status of the CpGs underlying the amplification primers. It was one of the most widely used technologies for DNA methylation analysis because it does require no expensive instrumentation and a large number of samples can be rapidly assessed. The main advantage of MSP is the high sensitivity, which enables the detection of one allele in the presence of a thousand-fold excess of the other (Herman et al., 1996). It permits the detection of methylation at high throughput, but this is limited to the primer binding sites. The detection limit can be further decreased by fluorescent labeling of one of the MSP-primers and analysis of the amplification product on a sequencer (Fig. 7.7) (Goessl et al., 2000). MSP is well-suited for the rapid and sensitive detection of methylation patterns at specific sequences, but it does not provide resolution at the individual nucleotide level, and heterogeneous methylation patterns at the primer binding sites can result in the failure of amplification. Furthermore, biased amplification leads to a more qualitative than quantitative result, which makes it difficult to distinguish different degrees of methylation at the target sites. An additional major drawback of MSP is the gel-based detection of the amplification products. Although it is acceptable for use in a research laboratory, diagnostics in a clinical setting require homogeneous high-throughput assays such as the fluorescent real-time MSP approaches described next (Cottrell and Laird, 2003).

Most real-time PCR-based methods use the same principle as the TaqMan assay (Holland et al., 1991). In addition to the two amplification primers, a third

TABLE 7.2 Single-Nucleotide Primer Extension Methods for Analysis of Individual CpG Positions After Bisulfite Treatment

Method	Label of Dideoxynucleotides	Separation	Detection Platform	Multiplexing Level	Comment	References
Methylation-sensitive-single-nucleotide primer extension (SNuPE)	Radioactive	Polyacrylamide gel electrophoresis	Phosphoimager	3–5 (primers differing in length)	Accurate but labor-intensive. Two reactions required for quantification	Gonzalgo and Jones (1997, 2002)
SNuPE immunoprecipitation reverse-phase high-performance liquid chromatography (HPLC)	None	Mass and hydrophobicity	Ion-pair reverse-phase HPLC	3 (non complementary tails at the 5′-end)	Medium-throughput	El-Maarri et al. (2002)
MethylQuant	None	Allele-specific primer extension with locked nucleic acid modified primers	Real-time thermocycler	No	Easy to implement, readily available instrumentation, high-throughput but two separate reactions for methylated and unmethylated allele	Thomassin et al. (2004)
GOOD assay	None	Mass	Matrix-assisted laser desorption (MALDI)/ionization mass spectrometry	3–5 (different masses)	High-throughput, simple mass spectrometric signatures	Tost et al. (2003a)
SNaPshot	Fluorescence	Capillary electrophoresis	Capillary electrophoresis	3 (noncomplementary tails at the 5′-end)	Medium-throughput, instrumentation available in many laboratories	Uhlmann et al. (2002) and Kaminsky et al. (2005)
Iplex	None	Mass	MALDI mass spectrometry	Up to 27	High-throughput	Ragoussis et al. (2006)

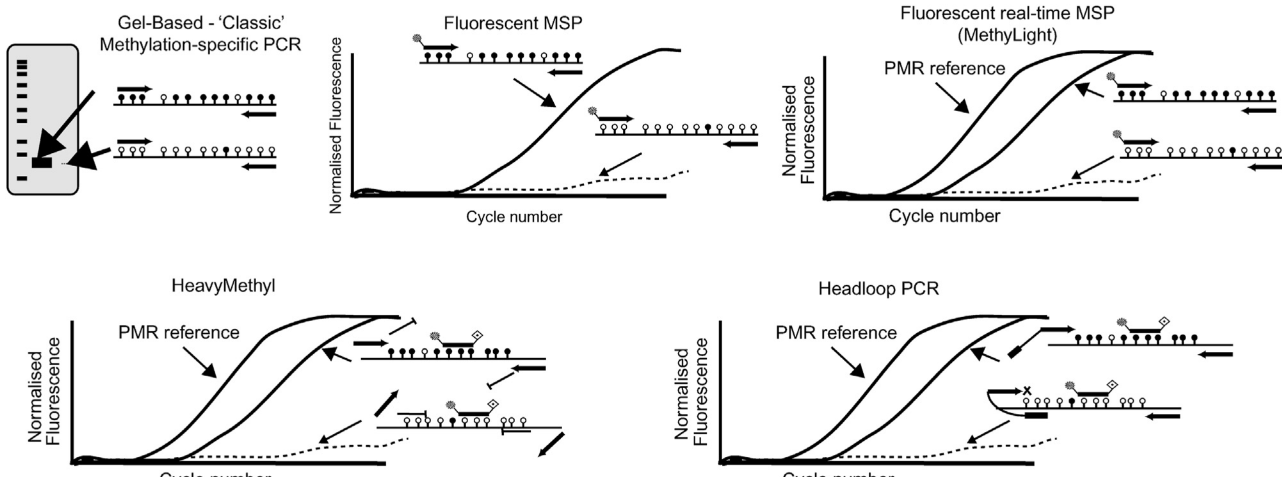


FIGURE 7.7 Real-time polymerase chain reaction (PCR) methods for the sensitive detection of DNA methylation in, for example, body fluids. For simplification, only amplification with a primer complementary to a completely methylated allele is shown. PMR indicates the percentage of methylated allele, a completely methylated DNA standard that is used to calculate the percentage of methylation contained within a sample. All methods use sodium bisulfite treatment before PCR amplification. CpGs are depicted as lollipops; former unmethylated CpGs are shown as open lollipops whereas filled ones correspond to former methylated CpGs. Although methylation is retained as a sequence difference after bisulfite treatment, lollipops are shown for easier differentiation of the alleles. Details of the different techniques are given in the text. *MSP*, methylation-specific polymerase chain reaction.

oligonucleotide called probe, which is dually labeled with a fluorescent reporter (e.g., FAM) and a quencher dye (e.g., TAMRA), hybridizes to a target sequence in the amplified region. The technology uses Förster resonance energy transfer: If the probe is fully complementary to the amplified sequence, the probe is cleaved by the $5' > 3'$ exonucleolytic activity of Taq polymerase during the extension phase of the amplification, which results in spatial separation of the two dyes and a fluorescent signal proportional to the amount of PCR product generated. Real-time quantitative PCR thereby enables the non-isotopic, rapid, and accurate quantitative analysis over a large dynamic range of several CpG positions in a small target region. The addition of a third probe that has to anneal correctly to the synthesized template improves sensitivity as well as specificity compared with conventional MSP, with the drawback that expensive hybridization probes are required. The simple, one-step procedure makes real-time methylation-specific PCRs rapid high-throughput assays for quantitative DNA methylation analysis that are robust and resistant to carryover contamination. These approaches provide information if molecules with a certain methylation pattern are present in the sample, such as conventional MSP, and also report on how many of them exist. Discrimination between methylated and unmethylated alleles is achieved at different levels of the primers and/or the hybridization probe (Eads et al., 2000a). Although in principle primers and probes could be designed for different combinations of methylated and unmethylated alleles, the most widely employed approaches such as MethyLight use primers and probes that are specific for the same methylation patterns, mostly completely methylated molecules. Heterogeneous methylation patterns that display large variations among consecutive CpGs complementary to primers or probe will therefore probably lead to failure of the assay or biased quantitative results.

7.3.6.2 MethyLight

MethyLight can detect a single hypermethylated allele against a background of 10,000 unmethylated alleles (Fig. 7.7) (Lo et al., 1999; Eads et al., 2000a). Absolute quantification of the number of molecules corresponding to the investigated pattern of methylation is calculated by the ratio between the gene of interest and a reference gene: for example, the *β-actin* gene. The percentage of fully methylated molecules (Fig. 7.7) [percentage of fully methylated reference (PMR)] is calculated by dividing the gene to reference ratio by the gene to completely methylated reference ratio (which is obtained by *Sss*I treatment of the normally unmethylated reference) and multiplying by 100 (Eads et al., 2000a). In a step-by-step technical evaluation, MethyLight displayed a high level of precision and reproducibility with an average variation of about 0.8 PMR (0.8%), with slightly larger variations induced by different bisulfite treatments (Ogino et al., 2006). The addition of an additional probe marked with a different fluorescent dye allows for the simultaneous detection of unconverted sequences that might coamplify with the bisulfite converted molecules, avoiding potential false-positive results (ConLight) (Rand et al., 2002).

In a variation of this approach, bisulfite-treated DNA was amplified independent of the methylation status in a first round of amplification (Fackler et al., 2004). This reaction

can be performed in multiplex, amplifying several target regions in a single reaction. In a nested PCR approach specifically methylated or unmethylated molecules are amplified in two separate reactions using the fluorescent real-time approach described earlier. This is in contrast to MethyLight, in which normally only the methylated molecules are interrogated (Weisenberger et al., 2006). Quantification of the percentage of methylation is performed by dividing the intensity of fluorescence corresponding to the methylated primer set to the added intensities of both sets, which yields results comparable to the reference gene approach. However, multiplex reactions might be difficult to set up on bisulfite-converted DNA with its low sequence diversity, and the equal amplification of methylated and unmethylated molecules (absence of a PCR bias) is a prerequisite for accurate quantification. Furthermore, partially methylated molecules are not taken into account, potentially overestimating the methylation content.

7.3.6.3 Quantitative Analysis of Methylated Alleles

QAMA is a variation of MethyLight that uses a TaqMan probe conjugated to a minor groove binder for discrimination at single-base resolution by forming hyperstabilized duplexes with complementary DNAs (Zeschnigk et al., 2004). Methylated and unmethylated alleles are simultaneously quantified using two probes modified with a quencher, a minor groove binder, and one of two fluorophores, VIC or FAM. Thus, amplification of bisulfite-treated DNA can be carried out with primers simultaneously amplifying formerly methylated and unmethylated alleles, and differentiation of the methylation status of alleles is achieved only at the probe level.

7.3.6.4 HeavyMethyl and Headloop Polymerase Chain Reaction

HeavyMethyl further increases sensitivity and specificity of real-time PCR-based assays for the analysis of DNA methylation using methylation-dependent blocking oligonucleotides (Fig. 7.7) (Cottrell et al., 2004). A similar approach was used successfully to detect genetic mutations diluted in an excess of normal DNA (Sun et al., 2002). In contrast to MethyLight, amplification primers are not specific for a certain methylation pattern but are positioned in sequence stretches containing no CpG positions. Only the fluorescent probe is specific, usually to consistently hypermethylated sequences. The increased specificity, and sensitivity is achieved through a second pair of non-extendable (3′-phosphorylated) oligonucleotides that hybridize specifically to a methylation pattern opposite the one investigated, usually the unmethylated sequence. The annealing sites of these oligonucleotides overlap with the target sequences for PCR amplification and thereby efficiently block any amplification of the bisulfite sequence corresponding to the undesired methylation pattern. HeavyMethyl was able to detect 25 pg of in vitro methylated DNA specifically in the background of 400 ng of unmethylated DNA (relative sensitivity up to 1:8000). As for MethyLight, a PMR standard is used to identify samples with negligible amounts of methylation. The use of four to five different oligonucleotides contributes significantly to the cost of the assay and design might be more complex compared with conventional MSP or MethyLight. However, HeavyMethyl shows the sensitivity and specificity necessary for clinical applications and might be a useful alternative for situations in which other real-time base methylation assays reach their limits. Of note, HeavyMethyl is the underlying technology for the Epi *pro*Colon test, analyzing methylation in the *SEPT9* gene (Church et al., 2014) and the *Epi pro*Lung (*SHOX2*) test (Ilse et al., 2014).

An alternative but similar approach uses amplification primers specific for a target after bisulfite treatment that carry a 5′-tail sequence complementary to a sequence present in the amplicon corresponding to a specific methylation pattern (Fig. 7.7) (Rand et al., 2005). After incorporation of the primer in the synthesized PCR product, the tail folds back onto the template, creating a secondary structure refractory to amplification. For example, if the tail is complementary to an unmethylated sequence, only methylated molecules are amplified. The amplification is monitored in real time with SYBR Green or by TaqMan probes. Sensitivity of the approach is similar to the others described earlier detecting a methylated allele in the presence of a 4000-fold excess of unmethylated ones. Therefore, this approach is well-suited to the selective amplification of tumor-derived methylated molecules in body fluids.

7.4 SINGLE-CELL DNA METHYLATION ANALYSIS

It has become clear that increasing the resolution of observing biological systems requires investigating individual cells as part of one cell type or according to their behavior as part of a group of cells in context. Working with single cells as part of a cell type requires the capture of individual cells. This is efficiently achieved with state-of-the-art cell isolation techniques. Initial -omics work on single cells focused on RNA analysis, and there has been an explosion of activity into the analysis of DNA methylation at the single-cell level. Sample preparation techniques for DNA methylation have been combined with multiplex assays using the single-cell restriction analysis of methylation assay, which applies methylation-sensitive restriction enzymes after single-cell isolation and lysis, and locus-specific qPCR on a Fluidigm Biomark system, which

allows the interrogation of 24 genomic positions in 48 single cells in one experiment (Lorthongpanich et al., 2013; Cheow et al., 2015). A single-cell DNA methylation analysis method that combines bisulfite conversion with Sanger sequencing and Agena's Epityper also allows cost-effective analysis of a larger number of single cells (Gravina et al., 2015).

Techniques combining single-cell analysis with NGS have become a major target. Apart from technical challenges, the main issue with single-cell analysis is the number of cells that need to be analyzed to obtain a representative picture of the biology and the amount of sequencing that can be afforded per cell. A balance needs to be struck between these two elements. Single-cell WGBS has been demonstrated (Smallwood et al., 2014). However, this approach recovers only 48.4% of all CpG positions of each cell analyzed. This means that far more cells need to be included to capture subpopulations and large oversampling is necessary. On the other hand, it allows non-CpG methylation to be captured. An alternative to preparing libraries for sequencing from individual cells is to generate pools of a small, defined number of cells, sequence each pool, and then use computational methods to deconvolute cell states as a result of the distortion in methylation detected between pools (Farlik et al., 2015). RRBS has been joined with single cells and is a good match in terms of genomic coverage, suitability with low input, and reduction in target size (Guo et al., 2013, 2015). It also allows quantitative analysis of differential DNA methylation (Wang et al., 2015).

7.5 CONCLUSIONS

DNA methylation analysis will undoubtedly have a key role in the diagnosis, prognostic assessment, and treatment of various diseases. A multitude of epigenetic markers have been discovered. However, for most complex diseases except perhaps cancer, we have only scratched the surface. A large increase in the number of identified epigenetic changes will be seen over the next few years, together with a rapid increase in publicly available genome-wide mammalian epigenome maps. Technology has been developed at a breathtaking speed and many companies now provide easy-to-implement assays and reproducible protocols for bisulfite treatment. With the technology available, research is shifting from gene-centered study to genome-wide analyses and cell-specific resolution even down to single cells.

One of the most important tasks is to validate known biomarkers in larger cohorts, because most studies have analyzed tens of samples instead of the hundreds or thousands required to confirm the utility of a biomarker definitely. The discovery of viable markers will crucially depend on access to correctly characterized and classified biological material. Another task is to translate the knowledge on epigenetic changes and their correlation with pathophysiologic parameters into clinical practice. The sensitive and specific detection of tumor-specific DNA methylation patterns at distal sites makes DNA methylation a biomarker of choice for the clinical management of patients, and predictive epigenetic biomarkers will allow a personalized treatment of disease based on the individual methylation profile.

REFERENCES

Absher, D.M., Li, X., Waite, L.L., Gibson, A., Roberts, K., Edberg, J., Chatham, W.W., Kimberly, R.P., 2013. Genome-wide DNA methylation analysis of systemic lupus erythematosus reveals persistent hypomethylation of interferon genes and compositional changes to $CD4^+$ T-cell populations. PLoS Genet. 9 (8), e1003678.

Adams, D., Altucci, L., Antonarakis, S.E., Ballesteros, J., Beck, S., Bird, A., Bock, C., Boehm, B., Campo, E., Caricasole, A., Dahl, F., Dermitzakis, E.T., Enver, T., Esteller, M., Estivill, X., Ferguson-Smith, A., Fitzgibbon, J., Flicek, P., Giehl, C., Graf, T., Grosveld, F., Guigo, R., Gut, I., Helin, K., Jarvius, J., Kuppers, R., Lehrach, H., Lengauer, T., Lernmark, A., Leslie, D., Loeffler, M., Macintyre, E., Mai, A., Martens, J.H., Minucci, S., Ouwehand, W.H., Pelicci, P.G., Pendeville, H., Porse, B., Rakyan, V., Reik, W., Schrappe, M., Schubeler, D., Seifert, M., Siebert, R., Simmons, D., Soranzo, N., Spicuglia, S., Stratton, M., Stunnenberg, H.G., Tanay, A., Torrents, D., Valencia, A., Vellenga, E., Vingron, M., Walter, J., Willcocks, S., 2012. BLUEPRINT to decode the epigenetic signature written in blood. Nat. Biotechnol. 30 (3), 224−226.

Adey, A., Morrison, H.G., Asan, Xun, X., Kitzman, J.O., Turner, E.H., Stackhouse, B., MacKenzie, A.P., Caruccio, N.C., Zhang, X., Shendure, J., 2010. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biol. 11 (12), R119.

Adey, A., Shendure, J., 2012. Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing. Genome Res. 22 (6), 1139−1143.

Adorján, P., Distler, J., Lipscher, E., Model, F., Müller, J., Pelet, C., Braun, A., Florl, A.R., Gütig, D., Grabs, G., Howe, A., Kursar, M., Lesche, R., Leu, E., Lewin, A., Maier, S., Müller, V., Otto, T., Scholz, C., Schulz, W.A., Seifert, H.H., Schwope, I., Ziebarth, H., Berlin, K., Piepenbrock, C., Olek, A., 2002. Tumour class prediction and discovery by microarray-based DNA methylation analysis. Nucleic Acids Res 30 (5), e21.

Adusumalli, S., Mohd Omar, M.F., Soong, R., Benoukraf, T., 2015. Methodological aspects of whole-genome bisulfite sequencing analysis. Brief. Bioinform. 16 (3), 369−379.

Akey, D.T., Akey, J.M., Zhang, K., Jin, L., 2002. Assaying DNA methylation based on high-throughput melting curve approaches. Genomics 80 (4), 376−384.

Alders, M., Bliek, J., vd Lip, K., vd Bogaard, R., Mannens, M., 2009. Determination of KCNQ1OT1 and H19 methylation levels in BWS and SRS patients using methylation-sensitive high-resolution melting analysis. Eur. J. Hum. Genet. 17 (4), 467−473.

Allum, F., Shao, X., Guenard, F., Simon, M.M., Busche, S., Caron, M., Lambourne, J., Lessard, J., Tandre, K., Hedman, A.K., Kwan, T.,

Ge, B., Ronnblom, L., McCarthy, M.I., Deloukas, P., Richmond, T., Burgess, D., Spector, T.D., Tchernof, A., Marceau, S., Lathrop, M., Vohl, M.C., Pastinen, T., Grundberg, E., Multiple Tissue Human Expression Resource Consortium, 2015. Characterization of functional methylomes by next-generation capture sequencing identifies novel disease-associated variants. Nat. Commun. 6, 7211.

Antequera, F., 2003. Structure, function and evolution of CpG island promoters. Cell. Mol. Life Sci. 60 (8), 1647–1658.

Balch, C., Yan, P., Craft, T., Young, S., Skalnik, D.G., Huang, T.H., Nephew, K.P., 2005. Antimitogenic and chemosensitizing effects of the methylation inhibitor zebularine in ovarian cancer. Mol. Cancer Ther. 4 (10), 1505–1514.

Balic, M., Pichler, M., Strutz, J., Heitzer, E., Ausch, C., Samonigg, H., Cote, R.J., Dandachi, N., 2009. High quality assessment of DNA methylation in archival tissues from colorectal cancer patients using quantitative high-resolution melting analysis. J. Mol. Diagn. 11 (2), 102–108.

Bastian, P.J., Palapattu, G.S., Lin, X., Yegnasubramanian, S., Mangold, L.A., Trock, B., Eisenberger, M.A., Partin, A.W., Nelson, W.G., 2005. Preoperative serum DNA GSTP1 CpG island hypermethylation and the risk of early prostate-specific antigen recurrence following radical prostatectomy. Clin. Cancer Res. 11 (11), 4037–4043.

Baylin, S.B., Ohm, J.E., 2006. Epigenetic gene silencing in cancer – a mechanism for early oncogenic pathway addiction? Nat. Rev. Cancer 6 (2), 107–116.

Berkhout, M., Nagtegaal, I.D., Cornelissen, S.J., Dekkers, M.M., van de Molengraft, F.J., Peters, W.H., Nagengast, F.M., van Krieken, J.H., Jeuken, J.W., 2007. Chromosomal and methylation alterations in sporadic and familial adenomatous polyposis-related duodenal carcinomas. Mod. Pathol. 20 (12), 1253–1262.

Bestor, T.H., Ingram, V.M., 1983. Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. Proc. Natl. Acad. Sci. U.S.A. 80 (18), 5559–5563.

Bhattacharyya, S., Yu, Y., Suzuki, M., Campbell, N., Mazdo, J., Vasanthakumar, A., Bhagat, T.D., Nischal, S., Christopeit, M., Parekh, S., Steidl, U., Godley, L., Maitra, A., Greally, J.M., Verma, A., 2013. Genome-wide hydroxymethylation tested using the HELP-GT assay shows redistribution in cancer. Nucleic Acids Res. 41 (16), e157.

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J.M., Delano, D., Zhang, L., Schroth, G.P., Gunderson, K.L., Fan, J.B., Shen, R., 2011. High density DNA methylation array with single CpG site resolution. Genomics 98 (4), 288–295.

Bibikova, M., Chudin, E., Wu, B., Zhou, L., Garcia, E.W., Liu, Y., Shin, S., Plaia, T.W., Auerbach, J.M., Arking, D.E., Gonzalez, R., Crook, J., Davidson, B., Schulz, T.C., Robins, A., Khanna, A., Sartipy, P., Hyllner, J., Vanguri, P., Savant-Bhonsale, S., Smith, A.K., Chakravarti, A., Maitra, A., Rao, M., Barker, D.L., Loring, J.F., Fan, J.B., 2006a. Human embryonic stem cells have a unique epigenetic signature. Genome Res. 16 (9), 1075–1083.

Bibikova, M., Le, J., Barnes, B., Saedinia-Melnyk, S., Zhou, L., Shen, R., Gunderson, K.L., 2009. Genome-wide DNA methylation profiling using Infinium® assay. Epigenomics 1 (1), 177–200.

Bibikova, M., Lin, Z., Zhou, L., Chudin, E., Garcia, E.W., Wu, B., Doucet, D., Thomas, N.J., Wang, Y., Vollmer, E., Goldmann, T., Seifart, C., Jiang, W., Barker, D.L., Chee, M.S., Floros, J., Fan, J.B., 2006b. High-throughput DNA methylation profiling using universal bead arrays. Genome Res. 16 (3), 383–393.

Bird, A., 2002. DNA methylation patterns and epigenetic memory. Genes Dev. 16 (1), 6–21.

Bird, A.P., Southern, E.M., 1978. Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from *Xenopus laevis*. J. Mol. Biol. 118 (1), 27–47.

Bjornsson, H.T., Brown, L.J., Fallin, M.D., Rongione, M.A., Bibikova, M., Wickham, E., Fan, J.B., Feinberg, A.P., 2007. Epigenetic specificity of loss of imprinting of the IGF2 gene in Wilms tumors. J. Natl. Cancer Inst. 99 (16), 1270–1273.

Bock, C., Tomazou, E.M., Brinkman, A.B., Muller, F., Simmer, F., Gu, H., Jager, N., Gnirke, A., Stunnenberg, H.G., Meissner, A., 2010. Quantitative comparison of genome-wide DNA methylation mapping technologies. Nat. Biotechnol. 28 (10), 1106–1114.

Bollati, V., Baccarelli, A., Hou, L., Bonzini, M., Fustinoni, S., Cavallo, D., Byun, H.M., Jiang, J., Marinelli, B., Pesatori, A.C., Bertazzi, P.A., Yang, A.S., 2007. Changes in DNA methylation patterns in subjects exposed to low-dose benzene. Cancer Res. 67 (3), 876–880.

Bönsch, D., Lenz, B., Reulbach, U., Kornhuber, J., Bleich, S., 2004. Homocysteine associated genomic DNA hypermethylation in patients with chronic alcoholism. J. Neural Transm. 111 (12), 1611–1616.

Bormann Chung, C.A., Boyd, V.L., McKernan, K.J., Fu, Y., Monighetti, C., Peckham, H.E., Barker, M., 2010. Whole methylome analysis by ultra-deep sequencing using two-base encoding. PLoS One 5 (2), e9320.

Borno, S.T., Fischer, A., Kerick, M., Falth, M., Laible, M., Brase, J.C., Kuner, R., Dahl, A., Grimm, C., Sayanjali, B., Isau, M., Rohr, C., Wunderlich, A., Timmermann, B., Claus, R., Plass, C., Graefen, M., Simon, R., Demichelis, F., Rubin, M.A., Sauter, G., Schlomm, T., Sultmann, H., Lehrach, H., Schweiger, M.R., 2012. Genome-wide DNA methylation events in TMPRSS2-ERG fusion-negative prostate cancers implicate an EZH2-dependent mechanism with miR-26a hypermethylation. Cancer Discov. 2 (11), 1024–1035.

Brandes, A.A., Tosoni, A., Cavallo, G., Reni, M., Franceschi, E., Bonaldi, L., Bertorelle, R., Gardiman, M., Ghimenton, C., Iuzzolino, P., Pession, A., Blatt, V., Ermani, M., 2006. Correlations between O6-methylguanine DNA methyltransferase promoter methylation status, 1p and 19q deletions, and response to temozolomide in anaplastic and recurrent oligodendroglioma: a prospective GICNO study. J. Clin. Oncol. 24 (29), 4746–4753.

Brena, R.M., Auer, H., Kornacker, K., Hackanson, B., Raval, A., Byrd, J.C., Plass, C., 2006. Accurate quantification of DNA methylation using combined bisulfite restriction analysis coupled with the Agilent 2100 Bioanalyzer platform. Nucleic Acids Res. 34 (3), e17.

Brinkman, A.B., Simmer, F., Ma, K., Kaan, A., Zhu, J., Stunnenberg, H.G., 2010. Whole-genome DNA methylation profiling using MethylCap-seq. Methods 52 (3), 232–236.

Brouilette, S., Kuersten, S., Mein, C., Bozek, M., Terry, A., Dias, K.R., Bhaw-Rosun, L., Shintani, Y., Coppen, S., Ikebe, C., Sawhney, V., Campbell, N., Kaneko, M., Tano, N., Ishida, H., Suzuki, K., Yashiro, K., 2012. A simple and novel method for RNA-seq library preparation of single cell cDNA analysis by hyperactive Tn5 transposase. Dev. Dyn. 241 (10), 1584–1590.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., Greenleaf, W.J., 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods 10 (12), 1213–1218.

Bundo, M., Sunaga, F., Ueda, J., Kasai, K., Kato, T., Iwamoto, K., 2012. A systematic evaluation of whole genome amplification of bisulfite-modified DNA. Clin. Epigenet. 4 (1), 22.

Burri, N., Chaubert, P., 1999. Complex methylation patterns analyzed by single-strand conformation polymorphism. Biotechniques 26 (2), 232–234.

Campan, M., Moffitt, M., Houshdaran, S., Shen, H., Widschwendter, M., Daxenbichler, G., Long, T., Marth, C., Laird-Offringa, I.A., Press, M.F., Dubeau, L., Siegmund, K.D., Wu, A.H., Groshen, S., Chandavarkar, U., Roman, L.D., Berchuck, A., Pearce, C.L., Laird, P.W., 2011. Genome-scale screen for DNA methylation-based detection markers for ovarian cancer. PLoS One 6 (12), e28141.

Castillo-Fernandez, J.E., Spector, T.D., Bell, J.T., 2014. Epigenetics of discordant monozygotic twins: implications for disease. Genome Med. 6 (7), 60.

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I.A., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., Correa, M., Da Silva, C., Just, J., Falentin, C., Koh, C.S., Le Clainche, I., Bernard, M., Bento, P., Noel, B., Labadie, K., Alberti, A., Charles, M., Arnaud, D., Guo, H., Daviaud, C., Alamery, S., Jabbari, K., Zhao, M., Edger, P.P., Chelaifa, H., Tack, D., Lassalle, G., Mestiri, I., Schnel, N., Le Paslier, M.C., Fan, G., Renault, V., Bayer, P.E., Golicz, A.A., Manoli, S., Lee, T.H., Thi, V.H., Chalabi, S., Hu, Q., Fan, C., Tollenaere, R., Lu, Y., Battail, C., Shen, J., Sidebottom, C.H., Wang, X., Canaguier, A., Chauveau, A., Berard, A., Deniot, G., Guan, M., Liu, Z., Sun, F., Lim, Y.P., Lyons, E., Town, C.D., Bancroft, I., Wang, X., Meng, J., Ma, J., Pires, J.C., King, G.J., Brunel, D., Delourme, R., Renard, M., Aury, J.M., Adams, K.L., Batley, J., Snowdon, R.J., Tost, J., Edwards, D., Zhou, Y., Hua, W., Sharpe, A.G., Paterson, A.H., Guan, C., Wincker, P., 2014. Plant genetics. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science 345 (6199), 950–953.

Chan, R.W., Jiang, P., Peng, X., Tam, L.S., Liao, G.J., Li, E.K., Wong, P.C., Sun, H., Chan, K.C., Chiu, R.W., Lo, Y.M., 2014. Plasma DNA aberrations in systemic lupus erythematosus revealed by genomic and methylomic sequencing. Proc. Natl. Acad. Sci. U.S.A. 111 (49), E5302–E5311.

Chandra, A., Ray, A., Senapati, S., Chatterjee, R., 2015. Genetic and epigenetic basis of psoriasis pathogenesis. Mol. Immunol. 64 (2), 313–323.

Chen, R.Z., Pettersson, U., Beard, C., Jackson-Grusby, L., Jaenisch, R., 1998. DNA hypomethylation leads to elevated mutation rates. Nature 395 (6697), 89–93.

Cheow, L.F., Quake, S.R., Burkholder, W.F., Messerschmidt, D.M., 2015. Multiplexed locus-specific analysis of DNA methylation in single cells. Nat. Protoc. 10 (4), 619–631.

Chouraki, V., Seshadri, S., 2014. Genetics of Alzheimer's disease. Adv. Genet. 87, 245–294.

Church, T.R., Wandell, M., Lofton-Day, C., Mongin, S.J., Burger, M., Payne, S.R., Castanos-Velez, E., Blumenstein, B.A., Rosch, T., Osborn, N., Snover, D., Day, R.W., Ransohoff, D.F., PRESEPT Clinical Study Steering Committee, Investigators and Study Team, 2014. Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. Gut 63 (2), 317–325.

Clark, S.J., Harrison, J., Paul, C.L., Frommer, M., 1994. High sensitivity mapping of methylated cytosines. Nucleic Acids Res. 22 (15), 2990–2997.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., Jacobsen, S.E., 2008. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature 452 (7184), 215–219.

Colella, S., Shen, L., Baggerly, K.A., Issa, J.P., Krahe, R., 2003. Sensitive and quantitative universal pyrosequencing methylation analysis of CpG sites. Biotechniques 35 (1), 146–150.

Coolen, M.W., Statham, A.L., Gardiner-Garden, M., Clark, S.J., 2007. Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements. Nucleic Acids Res. 35 (18), e119.

Costello, J.F., Fruhwald, M.C., Smiraglia, D.J., Rush, L.J., Robertson, G.P., Gao, X., Wright, F.A., Feramisco, J.D., Peltomaki, P., Lang, J.C., Schuller, D.E., Yu, L., Bloomfield, C.D., Caligiuri, M.A., Yates, A., Nishikawa, R., Su Huang, H., Petrelli, N.J., Zhang, X., O'Dorisio, M.S., Held, W.A., Cavenee, W.K., Plass, C., 2000. Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. Nat. Genet. 24 (2), 132–138.

Costello, J.F., Plass, C., Cavenee, W.K., 2002. Restriction landmark genome scanning. Methods Mol. Biol. 200, 53–70.

Cottrell, S.E., Distler, J., Goodman, N.S., Mooney, S.H., Kluth, A., Olek, A., Schwope, I., Tetzner, R., Ziebarth, H., Berlin, K., 2004. A real-time PCR assay for DNA-methylation using methylation-specific blockers. Nucleic Acids Res. 32 (1), e10.

Cottrell, S.E., Laird, P.W., 2003. Sensitive detection of DNA methylation. Ann. N.Y. Acad. Sci. 983, 120–130.

Crichton, J.H., Dunican, D.S., Maclennan, M., Meehan, R.R., Adams, I.R., 2014. Defending the genome from the enemy within: mechanisms of retrotransposon suppression in the mouse germline. Cell. Mol. Life Sci. 71 (9), 1581–1605.

Dai, W., Teodoridis, J.M., Zeller, C., Graham, J., Hersey, J., Flanagan, J.M., Stronach, E., Millan, D.W., Siddiqui, N., Paul, J., Brown, R., 2011. Systematic CpG islands methylation profiling of genes in the wnt pathway in epithelial ovarian cancer identifies biomarkers of progression-free survival. Clin. Cancer Res. 17 (12), 4052–4062.

Defossez, P.A., Stancheva, I., 2011. Biological functions of methyl-CpG-binding proteins. Prog. Mol. Biol. Transl. Sci. 101, 377–398.

Deneberg, S., Grövdal, M., Karimi, M., Jansson, M., Nahi, H., Corbacioglu, A., Gaidzik, V., Döhner, K., Paul, C., Ekström, T.J., Hellström-Lindberg, E., Lehmann, S., 2010. Gene-specific and global methylation patterns predict outcome in patients with acute myeloid leukemia. Leukemia 24 (5), 932–941.

Deng, D., Deng, G., Smith, M.F., Zhou, J., Xin, H., Powell, S.M., Lu, Y., 2002. Simultaneous detection of CpG methylation and single nucleotide polymorphism by denaturing high performance liquid chromatography. Nucleic Acids Res. 30 (3), e13.

Dey, B.K., Mueller, A.C., Dutta, A., 2014. Long non-coding RNAs as emerging regulators of differentiation, development, and disease. Transcription 5 (4), e944014.

Dick, K.J., Nelson, C.P., Tsaprouni, L., Sandling, J.K., Aissi, D., Wahl, S., Meduri, E., Morange, P.E., Gagnon, F., Grallert, H., Waldenberger, M., Peters, A., Erdmann, J., Hengstenberg, C., Cambien, F., Goodall, A.H., Ouwehand, W.H., Schunkert, H., Thompson, J.R., Spector, T.D., Gieger, C., Tregouet, D.A., Deloukas, P., Samani, N.J., 2014. DNA methylation and body-mass index: a genome-wide analysis. Lancet 383 (9933), 1990–1998.

Dikow, N., Nygren, A.O., Schouten, J.P., Hartmann, C., Kramer, N., Janssen, B., Zschocke, J., 2007. Quantification of the methylation

status of the PWS/AS imprinted region: comparison of two approaches based on bisulfite sequencing and methylation-sensitive MLPA. Mol. Cell Probes 21 (3), 208−215.

Down, T.A., Rakyan, V.K., Turner, D.J., Flicek, P., Li, H., Kulesha, E., Graf, S., Johnson, N., Herrero, J., Tomazou, E.M., Thorne, N.P., Backdahl, L., Herberth, M., Howe, K.L., Jackson, D.K., Miretti, M.M., Marioni, J.C., Birney, E., Hubbard, T.J., Durbin, R., Tavare, S., Beck, S., 2008. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. Nat. Biotechnol. 26 (7), 779−785.

Dupont, J.M., Tost, J., Jammes, H., Gut, I.G., 2004. De novo quantitative bisulfite sequencing using the pyrosequencing technology. Anal. Biochem. 333 (1), 119−127.

Eads, C.A., Danenberg, K.D., Kawakami, K., Saltz, L.B., Blake, C., Shibata, D., Danenberg, P.V., Laird, P.W., 2000a. MethyLight: a high-throughput assay to measure DNA methylation. Nucleic Acids Res. 28 (8), e32.

Eads, C.A., Lord, R.V., Kurumboor, S.K., Wickramasinghe, K., Skinner, M.L., Long, T.I., Peters, J.H., DeMeester, T.R., Danenberg, K.D., Danenberg, P.V., Laird, P.W., Skinner, K.A., 2000b. Fields of aberrant CpG island hypermethylation in Barrett's esophagus and associated adenocarcinoma. Cancer Res. 60 (18), 5021−5026.

Eden, A., Gaudet, F., Waghmare, A., Jaenisch, R., 2003. Chromosomal instability and tumors promoted by DNA hypomethylation. Science 300 (5618), 455.

Edwards, J.R., O'Donnell, A.H., Rollins, R.A., Peckham, H.E., Lee, C., Milekic, M.H., Chanrion, B., Fu, Y., Su, T., Hibshoosh, H., Gingrich, J.A., Haghighi, F., Nutter, R., Bestor, T.H., 2010. Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. Genome Res. 20 (7), 972−980.

Eggermann, T., Perez de Nanclares, G., Maher, E.R., Temple, I.K., Tumer, Z., Monk, D., Mackay, D.J., Gronskov, K., Riccio, A., Linglart, A., Netchine, I., 2015. Imprinting disorders: a group of congenital disorders with overlapping patterns of molecular changes affecting imprinted loci. Clin. Epigenet. 7, 123.

Ehrich, M., Nelson, M.R., Stanssens, P., Zabeau, M., Liloglou, T., Xinarianos, G., Cantor, C.R., Field, J.K., van den Boom, D., 2005. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proc. Natl. Acad. Sci. U.S.A. 102 (44), 15785−15790.

Ehrich, M., Turner, J., Gibbs, P., Lipton, L., Giovanneti, M., Cantor, C., van den Boom, D., 2008. Cytosine methylation profiling of cancer cell lines. Proc. Natl. Acad. Sci. U.S.A. 105 (12), 4844−4849.

Ehrich, M., Zoll, S., Sur, S., van den Boom, D., 2007. A new method for accurate assessment of DNA quality after bisulfite treatment. Nucleic Acids Res. 35 (5), e29.

Ehrlich, M., 2002. DNA methylation in cancer: too much, but also too little. Oncogene 21 (35), 5400−5413.

Eick, D., Fritz, H.J., Doerfler, W., 1983. Quantitative determination of 5-methylcytosine in DNA by reverse-phase high-performance liquid chromatography. Anal. Biochem. 135 (1), 165−171.

El-Maarri, O., Herbiniaux, U., Walter, J., Oldenburg, J., 2002. A rapid, quantitative, non-radioactive bisulfite-SNuPE- IP RP HPLC assay for methylation analysis at specific CpG sites. Nucleic Acids Res. 30 (6), e25.

Esteller, M., Hamilton, S.R., Burger, P.C., Baylin, S.B., Herman, J.G., 1999. Inactivation of the DNA repair gene O6-methylguanine-DNA methyltransferase by promoter hypermethylation is a common event in primary human neoplasia. Cancer Res. 59 (4), 793−797.

Etcheverry, A., Aubry, M., de Tayrac, M., Vauleon, E., Boniface, R., Guenot, F., Saikali, S., Hamlat, A., Riffaud, L., Menei, P., Quillien, V., Mosser, J., 2010. DNA methylation in glioblastoma: impact on gene expression and clinical outcome. BMC Genomics 11, 701.

Fachal, L., Dunning, A.M., 2015. From candidate gene studies to GWAS and post-GWAS analyses in breast cancer. Curr. Opin. Genet. Dev. 30, 32−41.

Fackler, M.J., McVeigh, M., Mehrotra, J., Blum, M.A., Lange, J., Lapides, A., Garrett, E., Argani, P., Sukumar, S., 2004. Quantitative multiplex methylation-specific PCR assay for the detection of promoter hypermethylation in multiple genes in breast cancer. Cancer Res. 64 (13), 4442−4452.

Fackler, M.J., Umbricht, C.B., Williams, D., Argani, P., Cruz, L.-A., Merino, V.F., Teo, W.W., Zhang, Z., Huang, P., Visvananthan, K., Marks, J., Ethier, S., Gray, J.W., Wolff, A.C., Cope, L.M., Sukumar, S., 2011. Genome-wide methylation analysis identifies genes specific to breast cancer hormone receptor status and risk of recurrence. Cancer Res. 71 (19), 6195−6207.

Farlik, M., Sheffield, N.C., Nuzzo, A., Datlinger, P., Schonegger, A., Klughammer, J., Bock, C., 2015. Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. Cell Rep. 10 (8), 1386−1397.

Fazzari, M.J., Greally, J.M., 2004. Epigenomics: beyond CpG islands. Nat. Rev. Genet. 5 (6), 446−455.

Feber, A., Guilhamon, P., Lechner, M., Fenton, T., Wilson, G.A., Thirlwell, C., Morris, T.J., Flanagan, A.M., Teschendorff, A.E., Kelly, J.D., Beck, S., 2014. Using high-density DNA methylation arrays to profile copy number alterations. Genome Biol. 15 (2), R30.

Feinberg, A.P., Ohlsson, R., Henikoff, S., 2006. The epigenetic progenitor origin of human cancer. Nat. Rev. Genet. 7 (1), 21−33.

Feinberg, A.P., Vogelstein, B., 1983. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature 301 (5895), 89−92.

Fernandez, A.F., Assenov, Y., Martin-Subero, J.I., Balint, B., Siebert, R., Taniguchi, H., Yamamoto, H., Hidalgo, M., Tan, A.C., Galm, O., Ferrer, I., Sanchez-Cespedes, M., Villanueva, A., Carmona, J., Sanchez-Mut, J.V., Berdasco, M., Moreno, V., Capella, G., Monk, D., Ballestar, E., Ropero, S., Martinez, R., Sanchez-Carbayo, M., Prosper, F., Agirre, X., Fraga, M.F., Grana, O., Perez-Jurado, L., Mora, J., Puig, S., Prat, J., Badimon, L., Puca, A.A., Meltzer, S.J., Lengauer, T., Bridgewater, J., Bock, C., Esteller, M., 2012. A DNA methylation fingerprint of 1628 human samples. Genome Res. 22 (2), 407−419.

Figueroa, M.E., Lugthart, S., Li, Y., Erpelinck-Verschueren, C., Deng, X., Christos, P.J., Schifano, E., Booth, J., van Putten, W., Skrabanek, L., Campagne, F., Mazumdar, M., Greally, J.M., Valk, P.J.M., Löwenberg, B., Delwel, R., Melnick, A., 2010. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. Cancer Cell 17 (1), 13−27.

Fouse, S.D., Nagarajan, R.O., Costello, J.F., 2010. Genome-scale DNA methylation analysis. Epigenomics 2 (1), 105−117.

Fraga, M.F., Ballestar, E., Montoya, G., Taysavang, P., Wade, P.A., Esteller, M., 2003. The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. Nucleic Acids Res. 31 (6), 1765−1774.

Fraga, M.F., Ballestar, E., Paz, M.F., Ropero, S., Setien, F., Ballestar, M.L., Heine-Suner, D., Cigudosa, J.C., Urioste, M., Benitez, J., Boix-Chornet, M., Sanchez-Aguilera, A., Ling, C., Carlsson, E., Poulsen, P., Vaag, A., Stephan, Z., Spector, T.D., Wu, Y.Z., Plass, C., Esteller, M., 2005. Epigenetic differences arise during the lifetime of monozygotic twins. Proc. Natl. Acad. Sci. U.S.A. 102 (30), 10604–10609.

Fraga, M.F., Uriol, E., Borja Diego, L., Berdasco, M., Esteller, M., Canal, M.J., Rodriguez, R., 2002. High-performance capillary electrophoretic method for the quantification of 5-methyl 2′-deoxycytidine in genomic DNA: application to plant, animal and human cancer tissues. Electrophoresis 23 (11), 1677–1681.

Frigola, J., Song, J., Stirzaker, C., Hinshelwood, R.A., Peinado, M.A., Clark, S.J., 2006. Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band. Nat. Genet. 38 (5), 540–549.

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., Paul, C.L., 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc. Natl. Acad. Sci. U.S.A. 89 (5), 1827–1831.

Geiman, T.M., Robertson, K.D., 2002. Chromatin remodeling, histone modifications, and DNA methylation—how does it all fit together? J. Cell Biochem. 87 (2), 117–125.

Gendrel, A.V., Heard, E., 2014. Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation. Annu. Rev. Cell Dev. Biol. 30, 561–580.

Globisch, D., Munzel, M., Muller, M., Michalakis, S., Wagner, M., Koch, S., Bruckl, T., Biel, M., Carell, T., 2010. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS One 5 (12), e15367.

Glossop, J.R., Emes, R.D., Nixon, N.B., Haworth, K.E., Packham, J.C., Dawes, P.T., Fryer, A.A., Mattey, D.L., Farrell, W.E., 2014. Genome-wide DNA methylation profiling in rheumatoid arthritis identifies disease-associated methylation changes that are distinct to individual T- and B-lymphocyte populations. Epigenetics 9 (9), 1228–1237.

Goelz, S.E., Vogelstein, B., Hamilton, S.R., Feinberg, A.P., 1985. Hypomethylation of DNA from benign and malignant human colon neoplasms. Science 228 (4696), 187–190.

Goessl, C., Krause, H., Muller, M., Heicappell, R., Schrader, M., Sachsinger, J., Miller, K., 2000. Fluorescent methylation-specific polymerase chain reaction for DNA-based detection of prostate cancer in bodily fluids. Cancer Res. 60 (21), 5941–5945.

Gonzalgo, M.L., Jones, P.A., 1997. Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). Nucleic Acids Res. 25 (12), 2529–2531.

Gonzalgo, M.L., Jones, P.A., 2002. Quantitative methylation analysis using methylation-sensitive single-nucleotide primer extension (Ms-SNuPE). Methods 27 (2), 128–133.

Gonzalgo, M.L., Liang, G., Spruck 3rd, C.H., Zingg, J.M., Rideout 3rd, W.M., Jones, P.A., 1997. Identification and characterization of differentially methylated regions of genomic DNA by methylation-sensitive arbitrarily primed PCR. Cancer Res. 57 (4), 594–599.

Gravina, S., Ganapathi, S., Vijg, J., 2015. Single-cell, locus-specific bisulfite sequencing (SLBS) for direct detection of epimutations in DNA methylation patterns. Nucleic Acids Res. 43 (14), e93.

Green, B.B., Marsit, C.J., 2015. Select prenatal environmental exposures and subsequent alterations of gene-specific and repetitive element DNA methylation in fetal tissues. Curr. Environ. Health Rep. 2 (2), 126–136.

Gries, J., Schumacher, D., Arand, J., Lutsik, P., Markelova, M.R., Fichtner, I., Walter, J., Sers, C., Tierling, S., 2013. Bi-PROF: bisulfite profiling of target regions using 454 GS FLX Titanium technology. Epigenetics 8 (7), 765–771.

Grunau, C., Clark, S.J., Rosenthal, A., 2001. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. Nucleic Acids Res. 29 (13), e65.

Grundberg, E., Meduri, E., Sandling, J.K., Hedman, A.K., Keildson, S., Buil, A., Busche, S., Yuan, W., Nisbet, J., Sekowska, M., Wilk, A., Barrett, A., Small, K.S., Ge, B., Caron, M., Shin, S.Y., Multiple Tissue Human Expression Resource Consortium, Lathrop, M., Dermitzakis, E.T., McCarthy, M.I., Spector, T.D., Bell, J.T., Deloukas, P., 2013. Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. Am. J. Hum. Genet. 93 (5), 876–890.

Guo, H., Zhu, P., Guo, F., Li, X., Wu, X., Fan, X., Wen, L., Tang, F., 2015. Profiling DNA methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing. Nat. Protoc. 10 (5), 645–659.

Guo, H., Zhu, P., Wu, X., Li, X., Wen, L., Tang, F., 2013. Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. Genome Res. 23 (12), 2126–2135.

Guo, J.U., Su, Y., Shin, J.H., Shin, J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., Zhu, H., Chang, Q., Gao, Y., Ming, G.L., Song, H., 2014. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nat. Neurosci. 17 (2), 215–222.

Hansen, K.D., Timp, W., Bravo, H.C., Sabunciyan, S., Langmead, B., McDonald, O.G., Wen, B., Wu, H., Liu, Y., Diep, D., Briem, E., Zhang, K., Irizarry, R.A., Feinberg, A.P., 2011. Increased methylation variation in epigenetic domains across cancer types. Nat. Genet. 43 (8), 768–775.

Hanson, M.A., Gluckman, P.D., 2014. Early developmental conditioning of later health and disease: physiology or pathophysiology? Physiol. Rev. 94 (4), 1027–1076.

Harris, R.A., Wang, T., Coarfa, C., Nagarajan, R.P., Hong, C., Downey, S.L., Johnson, B.E., Fouse, S.D., Delaney, A., Zhao, Y., Olshen, A., Ballinger, T., Zhou, X., Forsberg, K.J., Gu, J., Echipare, L., O'Geen, H., Lister, R., Pelizzola, M., Xi, Y., Epstein, C.B., Bernstein, B.E., Hawkins, R.D., Ren, B., Chung, W.Y., Gu, H., Bock, C., Gnirke, A., Zhang, M.Q., Haussler, D., Ecker, J.R., Li, W., Farnham, P.J., Waterland, R.A., Meissner, A., Marra, M.A., Hirst, M., Milosavljevic, A., Costello, J.F., 2010. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. Nat. Biotechnol. 28 (10), 1097–1105.

Hatada, I., Fukasawa, M., Kimura, M., Morita, S., Yamada, K., Yoshikawa, T., Yamanaka, S., Endo, C., Sakurada, A., Sato, M., Kondo, T., Horii, A., Ushijima, T., Sasaki, H., 2006. Genome-wide profiling of promoter methylation in human. Oncogene 25 (21), 3059–3064.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C.X., Zhang, K., He, C., Xu, G.L., 2011. Tet-mediated formation of 5-carboxylcytosine

and its excision by TDG in mammalian DNA. Science 333 (6047), 1303−1307.

Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Kallioniemi, O.P., Wilfond, B., Borg, A., Trent, J., Raffeld, M., Yakhini, Z., Ben-Dor, A., Dougherty, E., Kononen, J., Bubendorf, L., Fehrle, W., Pittaluga, S., Gruvberger, S., Loman, N., Johannsson, O., Olsson, H., Sauter, G., 2001. Gene-expression profiles in hereditary breast cancer. N. Engl. J. Med. 344 (8), 539−548.

Hegi, M.E., Diserens, A.C., Gorlia, T., Hamou, M.F., de Tribolet, N., Weller, M., Kros, J.M., Hainfellner, J.A., Mason, W., Mariani, L., Bromberg, J.E., Hau, P., Mirimanoff, R.O., Cairncross, J.G., Janzer, R.C., Stupp, R., 2005. MGMT gene silencing and benefit from temozolomide in glioblastoma. N. Engl. J. Med. 352 (10), 997−1003.

Herman, J.G., Graff, J.R., Myohanen, S., Nelkin, B.D., Baylin, S.B., 1996. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proc. Natl. Acad. Sci. U.S.A. 93 (18), 9821−9826.

Holland, P.M., Abramson, R.D., Watson, R., Gelfand, D.H., 1991. Detection of specific polymerase chain reaction product by utilizing the 5′-3′ exonuclease activity of Thermus aquaticus DNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 88 (16), 7276−7280.

How-Kit, A., Nielsen, H.M., Tost, J., 2012. DNA methylation based biomarkers: practical considerations and applications. Biochimie 94 (11), 2314−2337.

How-Kit, A., Dejeux, E., Dousset, B., Renault, V., Baudry, M., Terris, B., Tost, J., 2015. DNA methylation profiles distinguish different subtypes of gastroenteropancreatic neuroendocrine tumors. Epigenomics 7 (8), 1245−1258.

Hu, C.Y., Mohtat, D., Yu, Y., Ko, Y.A., Shenoy, N., Bhattacharya, S., Izquierdo, M.C., Park, A.S., Giricz, O., Vallumsetla, N., Gundabolu, K., Ware, K., Bhagat, T.D., Suzuki, M., Pullman, J., Liu, X.S., Greally, J.M., Susztak, K., Verma, A., 2014. Kidney cancer is characterized by aberrant methylation of tissue-specific enhancers that are prognostic for overall survival. Clin. Cancer Res. 20 (16), 4349−4360.

Huang, J., Renault, V., Sengenes, J., Touleimat, N., Michel, S., Lathrop, M., Tost, J., 2012. MeQA: a pipeline for MeDIP-seq data quality assessment and analysis. Bioinformatics 28 (4), 587−588.

Huang, T.H., Laux, D.E., Hamlin, B.C., Tran, P., Tran, H., Lubahn, D.B., 1997. Identification of DNA methylation markers for human breast carcinomas using the methylation-sensitive restriction fingerprinting technique. Cancer Res. 57 (6), 1030−1034.

Huang, T.H., Perry, M.R., Laux, D.E., 1999. Methylation profiling of CpG islands in human breast cancer cells. Hum. Mol. Genet. 8 (3), 459−470.

Hur, K., Cejas, P., Feliu, J., Moreno-Rubio, J., Burgos, E., Boland, C.R., Goel, A., 2014. Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of proto-oncogenes in human colorectal cancer metastasis. Gut 63 (4), 635−646.

Ilse, P., Biesterfeld, S., Pomjanski, N., Wrobel, C., Schramm, M., 2014. Analysis of SHOX2 methylation as an aid to cytology in lung cancer diagnosis. Cancer Genomics Proteomics 11 (5), 251−258.

Imperiale, T.F., Ransohoff, D.F., Itzkowitz, S.H., Levin, T.R., Lavin, P., Lidgard, G.P., Ahlquist, D.A., Berger, B.M., 2014. Multitarget stool DNA testing for colorectal-cancer screening. N. Engl. J. Med. 370 (14), 1287−1297.

Inamura, K., Yamauchi, M., Nishihara, R., Lochhead, P., Qian, Z.R., Kuchiba, A., Kim, S.A., Mima, K., Sukawa, Y., Jung, S., Zhang, X., Wu, K., Cho, E., Chan, A.T., Meyerhardt, J.A., Harris, C.C., Fuchs, C.S., Ogino, S., 2014. Tumor LINE-1 methylation level and microsatellite instability in relation to colorectal cancer prognosis. J. Natl. Cancer Inst. 106 (9).

Irizarry, R.A., Ladd-Acosta, C., Carvalho, B., Wu, H., Brandenburg, S.A., Jeddeloh, J.A., Wen, B., Feinberg, A.P., 2008. Comprehensive high-throughput arrays for relative methylation (CHARM). Genome Res. 18 (5), 780−790.

Issa, J.P., Ahuja, N., Toyota, M., Bronner, M.P., Brentnall, T.A., 2001. Accelerated age-related CpG island methylation in ulcerative colitis. Cancer Res. 61 (9), 3573−3577.

Ivanov, M., Kals, M., Kacevska, M., Metspalu, A., Ingelman-Sundberg, M., Milani, L., 2013. In-solution hybrid capture of bisulfite-converted DNA for targeted bisulfite sequencing of 174 ADME genes. Nucleic Acids Res. 41 (6), e72.

Jacinto, F.V., Esteller, M., 2007. MGMT hypermethylation: a prognostic foe, a predictive friend. DNA Repair 6 (8), 1155−1160.

Jenkins, T.G., Aston, K.I., Pflueger, C., Cairns, B.R., Carrell, D.T., 2014. Age-associated sperm DNA methylation alterations: possible implications in offspring disease susceptibility. PLoS Genet. 10 (7), e1004458.

Jones, P.A., Baylin, S.B., 2007. The epigenomics of cancer. Cell 128 (4), 683−692.

Jung, M., Kadam, S., Xiong, W., Rauch, T.A., Jin, S.G., Pfeifer, G.P., 2015. MIRA-seq for DNA methylation analysis of CpG islands. Epigenomics 7 (5), 695−706.

Jones, T.I., King, O.D., Homma, S., Chen, J.C., Beermann, M.L., Yan, C., Emerson Jr., C.P., Miller, J.B., Wagner, K.R., Jones, P.L., 2015. Individual epigenetic status of the pathogenic D4Z4 macrosatellite correlates with disease in facioscapulohumeral muscular dystrophy. Clin. Eig. 7 (1), 37.

Jurkowski, T.P., Ravichandran, M., Stepper, P., 2015. Synthetic epigenetics-towards intelligent control of epigenetic states and cell identity. Clin. Epigenet. 7 (1), 18.

Kaminsky, Z.A., Assadzadeh, A., Flanagan, J., Petronis, A., 2005. Single nucleotide extension technology for quantitative site-specific evaluation of metC/C in GC-rich regions. Nucleic Acids Res. 33 (10), e95.

Kantarjian, H., Issa, J.P., Rosenfeld, C.S., Bennett, J.M., Albitar, M., DiPersio, J., Klimek, V., Slack, J., de Castro, C., Ravandi, F., Helmer 3rd, R., Shen, L., Nimer, S.D., Leavitt, R., Raza, A., Saba, H., 2006. Decitabine improves patient outcomes in myelodysplastic syndromes: results of a phase III randomized study. Cancer 106 (8), 1794−1803.

Karatzas, P.S., Gazouli, M., Safioleas, M., Mantzaris, G.J., 2014. DNA methylation changes in inflammatory bowel disease. Ann. Gastroenterol. 27 (2), 125−132.

Karimi, M., Johansson, S., Stach, D., Corcoran, M., Grander, D., Schalling, M., Bakalkin, G., Lyko, F., Larsson, C., Ekstrom, T.J., 2006. LUMA (LUminometric Methylation Assay) − a high throughput method to the analysis of genomic DNA methylation. Exp. Cell Res. 312 (11), 1989−1995.

Kerkel, K., Spadola, A., Yuan, E., Kosek, J., Jiang, L., Hod, E., Li, K., Murty, V.V., Schupf, N., Vilain, E., Morris, M., Haghighi, F.,

Tycko, B., 2008. Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. Nat. Genet. 40 (7), 904–908.

Khulan, B., Thompson, R.F., Ye, K., Fazzari, M.J., Suzuki, M., Stasiek, E., Figueroa, M.E., Glass, J.L., Chen, Q., Montagna, C., Hatchwell, E., Selzer, R.R., Richmond, T.A., Green, R.D., Melnick, A., Greally, J.M., 2006. Comparative isoschizomer profiling of cytosine methylation: the HELP assay. Genome Res. 16 (8), 1046–1055.

Kobayashi, H., Sakurai, T., Imai, M., Takahashi, N., Fukuda, A., Yayoi, O., Sato, S., Nakabayashi, K., Hata, K., Sotomaru, Y., Suzuki, Y., Kono, T., 2012. Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. PLoS Genet. 8 (1), e1002440.

Kobayashi, H., Sakurai, T., Miura, F., Imai, M., Mochiduki, K., Yanagisawa, E., Sakashita, A., Wakai, T., Suzuki, Y., Ito, T., Matsui, Y., Kono, T., 2013. High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. Genome Res. 23 (4), 616–627.

Korbie, D., Lin, E., Wall, D., Nair, S.S., Stirzaker, C., Clark, S.J., Trau, M., 2015. Multiplex bisulfite PCR resequencing of clinical FFPE DNA. Clin. Epigenet. 7 (1), 28.

Kreck, B., Marnellos, G., Richter, J., Krueger, F., Siebert, R., Franke, A., 2012. B-SOLANA: an approach for the analysis of two-base encoding bisulfite sequencing data. Bioinformatics 28 (3), 428–429.

Kreck, B., Richter, J., Ammerpohl, O., Barann, M., Esser, D., Petersen, B.S., Vater, I., Murga Penas, E.M., Bormann Chung, C.A., Seisenberger, S., Lee Boyd, V., Smallwood, S., Drexler, H.G., Macleod, R.A., Hummel, M., Krueger, F., Hasler, R., Schreiber, S., Rosenstiel, P., Franke, A., Siebert, R., 2013. Base-pair resolution DNA methylome of the EBV-positive Endemic Burkitt lymphoma cell line DAUDI determined by SOLiD bisulfite-sequencing. Leukemia 27 (8), 1751–1753.

Kriaucionis, S., Heintz, N., 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science 324 (5929), 929–930.

Krishnakumar, R., Blelloch, R.H., 2013. Epigenetics of cellular reprogramming. Curr. Opin. Genet. Dev. 23 (5), 548–555.

Krueger, F., Andrews, S.R., 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27 (11), 1571–1572.

Kulis, M., Heath, S., Bibikova, M., Queiros, A.C., Navarro, A., Clot, G., Martinez-Trillos, A., Castellano, G., Brun-Heath, I., Pinyol, M., Barberan-Soler, S., Papasaikas, P., Jares, P., Bea, S., Rico, D., Ecker, S., Rubio, M., Royo, R., Ho, V., Klotzle, B., Hernandez, L., Conde, L., Lopez-Guerra, M., Colomer, D., Villamor, N., Aymerich, M., Rozman, M., Bayes, M., Gut, M., Gelpi, J.L., Orozco, M., Fan, J.B., Quesada, V., Puente, X.S., Pisano, D.G., Valencia, A., Lopez-Guillermo, A., Gut, I., Lopez-Otin, C., Campo, E., Martin-Subero, J.I., 2012. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nat. Genet. 44 (11), 1236–1242.

Ladd-Acosta, C., 2015. Epigenetic signatures as biomarkers of exposure. Curr. Environ. Health Rep. 2 (2), 117–125.

Ladd-Acosta, C., Pevsner, J., Sabunciyan, S., Yolken, R.H., Webster, M.J., Dinkins, T., Callinan, P.A., Fan, J.B., Potash, J.B., Feinberg, A.P., 2007. DNA methylation signatures within the human brain. Am. J. Hum. Genet. 81 (6), 1304–1315.

Laird, P.W., 2003. The power and the promise of DNA methylation markers. Nat. Rev. Cancer 3 (3), 253–266.

Laird, P.W., 2005. Cancer epigenetics. Hum. Mol. Genet 14 (Spec. No 1), R65–R76.

Le, T., Kim, K.P., Fan, G., Faull, K.F., 2011. A sensitive mass spectrometry method for simultaneous quantification of DNA methylation and hydroxymethylation levels in biological samples. Anal. Biochem. 412 (2), 203–209.

Lee, E.J., Luo, J., Wilson, J.M., Shi, H., 2013. Analyzing the cancer methylome through targeted bisulfite sequencing. Cancer Lett. 340 (2), 171–178.

Lee, E.J., Pei, L., Srivastava, G., Joshi, T., Kushwaha, G., Choi, J.H., Robertson, K.D., Wang, X., Colbourne, J.K., Zhang, L., Schroth, G.P., Xu, D., Zhang, K., Shi, H., 2011. Targeted bisulfite sequencing by solution hybrid selection and massively parallel sequencing. Nucleic Acids Res. 39 (19), e127.

Lewin, J., Schmitt, A.O., Adorjan, P., Hildmann, T., Piepenbrock, C., 2004. Quantitative DNA methylation analysis based on four-dye trace data from direct sequencing of PCR amplificates. Bioinformatics 20 (17), 3005–3012.

Li, D., Zhang, B., Xing, X., Wang, T., 2015. Combining MeDIP-seq and MRE-seq to investigate genome-wide CpG methylation. Methods 72, 29–40.

Li, Y., Zhu, J., Tian, G., Li, N., Li, Q., Ye, M., Zheng, H., Yu, J., Wu, H., Sun, J., Zhang, H., Chen, Q., Luo, R., Chen, M., He, Y., Jin, X., Zhang, Q., Yu, C., Zhou, G., Huang, Y., Cao, H., Zhou, X., Guo, S., Hu, X., Li, X., Kristiansen, K., Bolund, L., Xu, J., Wang, W., Yang, H., Wang, J., Li, R., Beck, S., Zhang, X., 2010. The DNA methylome of human peripheral blood mononuclear cells. PLoS Biol. 8 (11), e1000533.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., Ecker, J.R., 2008. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133 (3), 523–536.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A.H., Thomson, J.A., Ren, B., Ecker, J.R., 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462 (7271), 315–322.

Lister, R., Pelizzola, M., Kida, Y.S., Hawkins, R.D., Nery, J.R., Hon, G., Antosiewicz-Bourget, J., O'Malley, R., Castanon, R., Klugman, S., Downes, M., Yu, R., Stewart, R., Ren, B., Thomson, J.A., Evans, R.M., Ecker, J.R., 2011. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature 471 (7336), 68–73.

Liu, Z., Liu, S., Xie, Z., Blum, W., Perrotti, D., Paschka, P., Klisovic, R., Byrd, J., Chan, K.K., Marcucci, G., 2007. Characterization of in vitro and in vivo hypomethylating effects of decitabine in acute myeloid leukemia by a rapid, specific and sensitive LC-MS/MS method. Nucleic Acids Res. 35 (5), e31.

Lo, Y.M., Wong, I.H., Zhang, J., Tein, M.S., Ng, M.H., Hjelm, N.M., 1999. Quantitative analysis of aberrant p16 methylation using real-time quantitative methylation-specific polymerase chain reaction. Cancer Res. 59 (16), 3899–3903.

Lorthongpanich, C., Cheow, L.F., Balu, S., Quake, S.R., Knowles, B.B., Burkholder, W.F., Solter, D., Messerschmidt, D.M., 2013. Single-cell DNA-methylation analysis reveals epigenetic chimerism in preimplantation embryos. Science 341 (6150), 1110–1112.

Luo, S., Preuss, D., 2003. Strand-biased DNA methylation associated with centromeric regions in Arabidopsis. Proc. Natl. Acad. Sci. U.S.A. 100 (19), 11133−11138.

Lutsik, P., Feuerbach, L., Arand, J., Lengauer, T., Walter, J., Bock, C., 2011. BiQ Analyzer HT: locus-specific analysis of DNA methylation by high-throughput bisulfite sequencing. Nucleic Acids Res. 39 (web server issue), W551−W556.

Lyko, F., Foret, S., Kucharski, R., Wolf, S., Falckenhayn, C., Maleszka, R., 2010. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. PLoS Biol. 8 (11), e1000506.

Maier, S., Dahlstroem, C., Haefliger, C., Plum, A., Piepenbrock, C., 2005. Identifying DNA methylation biomarkers of cancer drug response. Am. J. Pharmacogenomics 5 (4), 223−232.

Maiti, A., Drohat, A.C., 2011. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. J. Biol. Chem. 286 (41), 35334−35338.

Marques-Rocha, J.L., Milagro, F.I., Mansego, M.L., Mourao, D.M., Martinez, J.A., Bressan, J., 2016. LINE-1 methylation is positively associated with healthier lifestyle but inversely related to body fat mass in healthy young individuals. Epigenetics 1−12.

Martinez, D., Pentinat, T., Ribo, S., Daviaud, C., Bloks, V.W., Cebria, J., Villalmanzo, N., Kalko, S.G., Ramon-Krauel, M., Diaz, R., Plosch, T., Tost, J., Jimenez-Chillaron, J.C., 2014. In utero undernutrition in male mice programs liver lipid metabolism in the second-generation offspring involving altered Lxra DNA methylation. Cell Metab. 19 (6), 941−951.

Martino, D., Joo, J.E., Sexton-Oates, A., Dang, T., Allen, K., Saffery, R., Prescott, S., 2014. Epigenome-wide association study reveals longitudinally stable DNA methylation differences in $CD4^+$ T cells from children with IgE-mediated food allergy. Epigenetics 9 (7), 998−1006.

Masser, D.R., Berg, A.S., Freeman, W.M., 2013. Focused, high accuracy 5-methylcytosine quantitation with base resolution by benchtop next-generation sequencing. Epigenetics Chromatin 6 (1), 33.

Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y., Turecki, G., Delaney, A., Varhol, R., Thiessen, N., Shchors, K., Heine, V.M., Rowitch, D.H., Xing, X., Fiore, C., Schillebeeckx, M., Jones, S.J., Haussler, D., Marra, M.A., Hirst, M., Wang, T., Costello, J.F., 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature 466 (7303), 253−257.

Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., Jaenisch, R., 2005. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. 33 (18), 5868−5877.

Melnikov, A.A., Gartenhaus, R.B., Levenson, A.S., Motchoulskaia, N.A., Levenson Chernokhvostov, V.V., 2005. MSRE-PCR for analysis of gene-specific DNA methylation. Nucleic Acids Res. 33 (10), e93.

Meng, H., Cao, Y., Qin, J., Song, X., Zhang, Q., Shi, Y., Cao, L., 2015. DNA methylation, its mediators and genome integrity. Int. J. Biol. Sci. 11 (5), 604−617.

Miceli-Richard, C., Wang-Renault, S.F., Boudaoud, S., Busato, F., Lallemand, C., Bethune, K., Belkhir, R., Nocturne, G., Mariette, X., Tost, J., 2015. Overlap between differentially methylated DNA regions in blood B lymphocytes and genetic at-risk loci in primary Sjogren's syndrome. Ann. Rheum. Dis. 75 (5), 933−940.

Michel, S., Busato, F., Genuneit, J., Pekkanen, J., Dalphin, J.C., Riedler, J., Mazaleyrat, N., Weber, J., Karvonen, A.M., Hirvonen, M.R., Braun-Fahrlander, C., Lauener, R., von Mutius, E., Kabesch, M., Tost, J., PASTURE Study Group, 2013. Farm exposure and time trends in early childhood may influence DNA methylation in genes related to asthma and allergy. Allergy 68 (3), 355−364.

Miller, O.J., Schnedl, W., Allen, J., Erlanger, B.F., 1974. 5-Methylcytosine localised in mammalian constitutive heterochromatin. Nature 251 (5476), 636−637.

Miura, F., Enomoto, Y., Dairiki, R., Ito, T., 2012. Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. Nucleic Acids Res. 40 (17), e136.

Miura, F., Ito, T., 2015. Highly sensitive targeted methylome sequencing by post-bisulfite adaptor tagging. DNA Res. 22 (1), 13−18.

Mohlke, K.L., Boehnke, M., 2015. Recent advances in understanding the genetic architecture of type 2 diabetes. Hum. Mol. Genet. 24 (R1), R85−R92.

Moran, S., Arribas, C., Esteller, M., 2015. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics 8.

Mund, C., Hackanson, B., Stresemann, C., Lubbert, M., Lyko, F., 2005. Characterization of DNA demethylation effects induced by 5-Aza-2′-deoxycytidine in patients with myelodysplastic syndrome. Cancer Res. 65 (16), 7086−7090.

Nair, S.S., Coolen, M.W., Stirzaker, C., Song, J.Z., Statham, A.L., Strbenac, D., Robinson, M.D., Clark, S.J., 2011. Comparison of methyl-DNA immunoprecipitation (MeDIP) and methyl-CpG binding domain (MBD) protein capture for genome-wide DNA methylation analysis reveal CpG sequence coverage bias. Epigenetics 6 (1), 34−44.

Nakagawachi, T., Soejima, H., Urano, T., Zhao, W., Higashimoto, K., Satoh, Y., Matsukura, S., Kudo, S., Kitajima, Y., Harada, H., Furukawa, K., Matsuzaki, H., Emi, M., Nakabeppu, Y., Miyazaki, K., Sekiguchi, M., Mukai, T., 2003. Silencing effect of CpG island hypermethylation and histone modifications on O6-methylguanine-DNA methyltransferase (MGMT) gene expression in human cancer. Oncogene 22 (55), 8835−8844.

Neidhart, M., Rethage, J., Kuchen, S., Kunzler, P., Crowl, R.M., Billingham, M.E., Gay, R.E., Gay, S., 2000. Retrotransposable L1 elements expressed in rheumatoid arthritis synovial tissue: association with genomic DNA hypomethylation and influence on gene expression. Arthritis Rheum. 43 (12), 2634−2647.

Newton, K., Jorgensen, N.M., Wallace, A.J., Buchanan, D.D., Lalloo, F., McMahon, R.F., Hill, J., Evans, D.G., 2014. Tumour MLH1 promoter region methylation testing is an effective prescreen for Lynch Syndrome (HNPCC). J. Med. Genet. 51 (12), 789−796.

Nones, K., Waddell, N., Song, S., Patch, A.M., Miller, D., Johns, A., Wu, J., Kassahn, K.S., Wood, D., Bailey, P., Fink, L., Manning, S., Christ, A.N., Nourse, C., Kazakoff, S., Taylor, D., Leonard, C., Chang, D.K., Jones, M.D., Thomas, M., Watson, C., Pinese, M., Cowley, M., Rooman, I., Pajic, M., APGI, Butturini, G., Malpaga, A., Corbo, V., Crippa, S., Falconi, M., Zamboni, G., Castelli, P., Lawlor, R.T., Gill, A.J., Scarpa, A., Pearson, J.V., Biankin, A.V., Grimmond, S.M., 2014. Genome-wide DNA methylation patterns in pancreatic ductal adenocarcinoma reveal epigenetic deregulation of SLIT-ROBO, ITGA2 and MET signaling. Int. J. Cancer 135 (5), 1110−1118.

Nygren, A.O., Ameziane, N., Duarte, H.M., Vijzelaar, R.N., Waisfisz, Q., Hess, C.J., Schouten, J.P., Errami, A., 2005. Methylation-specific MLPA (MS-MLPA): simultaneous detection of CpG methylation and copy number changes of up to 40 sequences. Nucleic Acids Res. 33 (14), e128.

Oakes, C.C., La Salle, S., Robaire, B., Trasler, J.M., 2006. Evaluation of a quantitative DNA methylation analysis technique using methylation-sensitive/dependent restriction enzymes and real-time PCR. Epigenetics 1 (3), 146−152.

Ogino, S., Kawasaki, T., Brahmandam, M., Cantor, M., Kirkner, G.J., Spiegelman, D., Makrigiorgos, G.M., Weisenberger, D.J., Laird, P.W., Loda, M., Fuchs, C.S., 2006. Precision and performance characteristics of bisulfite conversion and real-time PCR (MethyLight) for quantitative DNA methylation analysis. J. Mol. Diagn. 8 (2), 209−217.

Olkhov-Mitsel, E., Bapat, B., 2012. Strategies for discovery and validation of methylated and hydroxymethylated DNA biomarkers. Cancer Med. 1 (2), 237−260.

Paliwal, A., Temkin, A.M., Kerkel, K., Yale, A., Yotova, I., Drost, N., Lax, S., Nhan-Chang, C.L., Powell, C., Borczuk, A., Aviv, A., Wapner, R., Chen, X., Nagy, P.L., Schork, N., Do, C., Torkamani, A., Tycko, B., 2013. Comparative anatomy of chromosomal domains with imprinted and non-imprinted allele-specific DNA methylation. PLoS Genet. 9 (8), e1003622.

Palmisano, W.A., Divine, K.K., Saccomanno, G., Gilliland, F.D., Baylin, S.B., Herman, J.G., Belinsky, S.A., 2000. Predicting lung cancer by detecting aberrant promoter methylation in sputum. Cancer Res. 60 (21), 5954−5958.

Peat, J.R., Dean, W., Clark, S.J., Krueger, F., Smallwood, S.A., Ficz, G., Kim, J.K., Marioni, J.C., Hore, T.A., Reik, W., 2014. Genome-wide bisulfite sequencing in zygotes identifies demethylation targets and maps the contribution of TET3 oxidation. Cell Rep. 9 (6), 1990−2000.

Pelizzola, M., Koga, Y., Urban, A.E., Krauthammer, M., Weissman, S., Halaban, R., Molinaro, A.M., 2008. MEDME: an experimental and analytical methodology for the estimation of DNA methylation levels based on microarray derived MeDIP-enrichment. Genome Res. 18 (10), 1652−1659.

Pogribny, I., Yi, P., James, S.J., 1999. A sensitive new method for rapid detection of abnormal methylation patterns in global DNA and within CpG islands. Biochem. Biophys. Res. Commun. 262 (3), 624−628.

Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., Reik, W., 2010. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature 463 (7284), 1101−1105.

Priolo, M., Sparago, A., Mammi, C., Cerrato, F., Lagana, C., Riccio, A., 2008. MS-MLPA is a specific and sensitive technique for detecting all chromosome 11p15.5 imprinting defects of BWS and SRS in a single-tube experiment. Eur. J. Hum. Genet. 16 (5), 565−571.

Pröll, J., Födermayr, M., Wechselberger, C., Pammer, P., Sonnleitner, M., Zach, O., Lutz, D., 2006. Ultra-sensitive immunodetection of 5′methyl cytosine for DNA methylation analysis on oligonucleotide microarrays. DNA Res. 13 (1), 37−42.

Radpour, R., Kohler, C., Haghighi, M.M., Fan, A.X., Holzgreve, W., Zhong, X.Y., 2009. Methylation profiles of 22 candidate genes in breast cancer using high-throughput MALDI-TOF mass array. Oncogene 28 (33), 2969−2978.

Ragoussis, J., Elvidge, G.P., Kaur, K., Colella, S., 2006. Matrix-assisted laser desorption/ionisation, time-of-flight mass spectrometry in genomics research. PLoS Genet. 2 (7), e100.

Rakyan, V.K., Down, T.A., Thorne, N.P., Flicek, P., Kulesha, E., Graf, S., Tomazou, E.M., Backdahl, L., Johnson, N., Herberth, M., Howe, K.L., Jackson, D.K., Miretti, M.M., Fiegler, H., Marioni, J.C., Birney, E., Hubbard, T.J., Carter, N.P., Tavare, S., Beck, S., 2008. An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). Genome Res. 18 (9), 1518−1529.

Rakyan, V.K., Hildmann, T., Novik, K.L., Lewin, J., Tost, J., Cox, A.V., Andrews, T.D., Howe, K.L., Otto, T., Olek, A., Fischer, J., Gut, I.G., Berlin, K., Beck, S., 2004. DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. PLoS Biol. 2 (12), e405.

Rand, K., Qu, W., Ho, T., Clark, S.J., Molloy, P., 2002. Conversion-specific detection of DNA methylation using real-time polymerase chain reaction (ConLight-MSP) to avoid false positives. Methods 27 (2), 114−120.

Rand, K.N., Ho, T., Qu, W., Mitchell, S.M., White, R., Clark, S.J., Molloy, P.L., 2005. Headloop suppression PCR and its application to selective amplification of methylated DNA sequences. Nucleic Acids Res. 33 (14), e127.

Rauch, T., Li, H., Wu, X., Pfeifer, G.P., 2006. MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells. Cancer Res. 66 (16), 7939−7947.

Rauch, T., Pfeifer, G.P., 2005. Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer. Lab. Invest. 85 (9), 1172−1180.

Reik, W., Dean, W., Walter, J., 2001. Epigenetic reprogramming in mammalian development. Science 293 (5532), 1089−1093.

Reik, W., Walter, J., 2001. Genomic imprinting: parental influence on the genome. Nat. Rev. Genet. 2 (1), 21−32.

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., Amin, V., Whitaker, J.W., Schultz, M.D., Ward, L.D., Sarkar, A., Quon, G., Sandstrom, R.S., Eaton, M.L., Wu, Y.C., Pfenning, A.R., Wang, X., Claussnitzer, M., Liu, Y., Coarfa, C., Harris, R.A., Shoresh, N., Epstein, C.B., Gjoneska, E., Leung, D., Xie, W., Hawkins, R.D., Lister, R., Hong, C., Gascard, P., Mungall, A.J., Moore, R., Chuah, E., Tam, A., Canfield, T.K., Hansen, R.S., Kaul, R., Sabo, P.J., Bansal, M.S., Carles, A., Dixon, J.R., Farh, K.H., Feizi, S., Karlic, R., Kim, A.R., Kulkarni, A., Li, D., Lowdon, R., Elliott, G., Mercer, T.R., Neph, S.J., Onuchic, V., Polak, P., Rajagopal, N., Ray, P., Sallari, R.C., Siebenthall, K.T., Sinnott-Armstrong, N.A., Stevens, M., Thurman, R.E., Wu, J., Zhang, B., Zhou, X., Beaudet, A.E., Boyer, L.A., De Jager, P.L., Farnham, P.J., Fisher, S.J., Haussler, D., Jones, S.J., Li, W., Marra, M.A., McManus, M.T., Sunyaev, S., Thomson, J.A., Tlsty, T.D., Tsai, L.H., Wang, W., Waterland, R.A., Zhang, M.Q., Chadwick, L.H., Bernstein, B.E., Costello, J.F., Ecker, J.R., Hirst, M., Meissner, A., Milosavljevic, A., Ren, B., Stamatoyannopoulos, J.A., Wang, T., Kellis, M., 2015. Integrative analysis of 111 reference human epigenomes. Nature 518 (7539), 317−330.

Robinson, M.D., Stirzaker, C., Statham, A.L., Coolen, M.W., Song, J.Z., Nair, S.S., Strbenac, D., Speed, T.P., Clark, S.J., 2010. Evaluation of

affinity-based genome-wide DNA methylation data: effects of CpG density, amplification bias, and copy number variation. Genome Res. 20 (12), 1719−1729.

Roessler, J., Ammerpohl, O., Gutwein, J., Hasemeier, B., Anwar, S.L., Kreipe, H., Lehmann, U., 2012. Quantitative cross-validation and content analysis of the 450k DNA methylation array from Illumina, Inc. BMC Res. Notes 5, 210.

Rollins, R.A., Haghighi, F., Edwards, J.R., Das, R., Zhang, M.Q., Ju, J., Bestor, T.H., 2006. Large-scale structure of genomic methylation patterns. Genome Res. 16 (2), 157−163.

Ronaghi, M., Uhlen, M., Nyren, P., 1998. A sequencing method based on real-time pyrophosphate. Science 281 (5375), 363−365.

Ronn, T., Ling, C., 2015. DNA methylation as a diagnostic and therapeutic target in the battle against Type 2 diabetes. Epigenomics 7 (3), 451−460.

Rosenfeld, N., Aharonov, R., Meiri, E., Rosenwald, S., Spector, Y., Zepeniuk, M., Benjamin, H., Shabes, N., Tabak, S., Levy, A., Lebanony, D., Goren, Y., Silberschein, E., Targan, N., Ben-Ari, A., Gilad, S., Sion-Vardy, N., Tobar, A., Feinmesser, M., Kharenko, O., Nativ, O., Nass, D., Perelman, M., Yosepovich, A., Shalmon, B., Polak-Charcon, S., Fridman, E., Avniel, A., Bentwich, I., Bentwich, Z., Cohen, D., Chajut, A., Barshack, I., 2008. MicroRNAs accurately identify cancer tissue origin. Nat. Biotechnol. 26 (4), 462−469.

Ruzov, A., Tsenkina, Y., Serio, A., Dudnakova, T., Fletcher, J., Bai, Y., Chebotareva, T., Pells, S., Hannoun, Z., Sullivan, G., Chandran, S., Hay, D.C., Bradley, M., Wilmut, I., De Sousa, P., 2011. Lineage-specific distribution of high levels of genomic 5-hydroxy-methylcytosine in mammalian development. Cell Res. 21 (9), 1332−1342.

Sadri, R., Hornsby, P.J., 1996. Rapid analysis of DNA methylation using new restriction enzyme sites created by bisulfite modification. Nucleic Acids Res. 24 (24), 5058−5059.

Sandoval, J., Heyn, H., Moran, S., Serra-Musach, J., Pujana, M.A., Bibikova, M., Esteller, M., 2011. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics 6 (6), 692−702.

Sanli, I., Feil, R., 2015. Chromatin mechanisms in the developmental control of imprinted gene expression. Int. J. Biochem. Cell Biol. 67, 139−147.

Scesnaite, A., Jarmalaite, S., Mutanen, P., Anttila, S., Nyberg, F., Benhamou, S., Boffetta, P., Husgafvel-Pursiainen, K., 2012. Similar DNA methylation pattern in lung tumours from smokers and never-smokers with second-hand tobacco smoke exposure. Mutagenesis 27 (4), 423−429.

Schatz, P., Dietrich, D., Schuster, M., 2004. Rapid analysis of CpG methylation patterns using RNase T1 cleavage and MALDI-TOF. Nucleic Acids Res. 32 (21), e167.

Schmidl, C., Rendeiro, A.F., Sheffield, N.C., Bock, C., 2015. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nat. Methods 12.

Schubeler, D., 2015. Function and information content of DNA methylation. Nature 517 (7534), 321−326.

Schwarzenbach, H., Nishida, N., Calin, G.A., Pantel, K., 2014. Clinical relevance of circulating cell-free microRNAs in cancer. Nat. Rev. Clin. Oncol. 11 (3), 145−156.

Scott, R.H., Douglas, J., Baskcomb, L., Nygren, A.O., Birch, J.M., Cole, T.R., Cormier-Daire, V., Eastwood, D.M., Garcia-Minaur, S., Lupunzina, P., Tatton-Brown, K., Bliek, J., Maher, E.R., Rahman, N., 2008. Methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA) robustly detects and distinguishes 11p15 abnormalities associated with overgrowth and growth retardation. J. Med. Genet. 45 (2), 106−113.

Selker, E.U., Tountas, N.A., Cross, S.H., Margolin, B.S., Murphy, J.G., Bird, A.P., Freitag, M., 2003. The methylated component of the Neurospora crassa genome. Nature 422 (6934), 893−897.

Serre, D., Lee, B.H., Ting, A.H., 2010. MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. Nucleic Acids Res. 38 (2), 391−399.

Shapiro, R., DiFate, V., Welcher, M., 1974. Deamination of cytosine derivatives by bisulfite. Mechanism of the reaction. J. Am. Chem. Soc. 96 (3), 906−912.

Shen, L., Kondo, Y., Guo, Y., Zhang, J., Zhang, L., Ahmed, S., Shu, J., Chen, X., Waterland, R.A., Issa, J.P., 2007. Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. PLoS Genet. 3 (10), 2023−2036.

Shirane, K., Toh, H., Kobayashi, H., Miura, F., Chiba, H., Ito, T., Kono, T., Sasaki, H., 2013. Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. PLoS Genet. 9 (4), e1003439.

Singer-Sam, J., LeBon, J.M., Tanguay, R.L., Riggs, A.D., 1990. A quantitative HpaII-PCR assay to measure methylation of DNA from a small number of cells. Nucleic Acids Res. 18 (3), 687.

Smallwood, S.A., Lee, H.J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S.R., Stegle, O., Reik, W., Kelsey, G., 2014. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat. Methods 11 (8), 817−820.

Stefansson, O.A., Moran, S., Gomez, A., Sayols, S., Arribas-Jorba, C., Sandoval, J., Hilmarsdottir, H., Olafsdottir, E., Tryggvadottir, L., Jonasson, J.G., Eyfjord, J., Esteller, M., 2015. A DNA methylation-based definition of biologically distinct breast cancer subtypes. Mol. Oncol. 9 (3), 555−568.

Stransky, N., Vallot, C., Reyal, F., Bernard-Pierrot, I., de Medina, S.G., Segraves, R., de Rycke, Y., Elvin, P., Cassidy, A., Spraggon, C., Graham, A., Southgate, J., Asselain, B., Allory, Y., Abbou, C.C., Albertson, D.G., Thiery, J.P., Chopin, D.K., Pinkel, D., Radvanyi, F., 2006. Regional copy number-independent deregulation of transcription in cancer. Nat. Genet. 38 (12), 1386−1396.

Stresemann, C., Brueckner, B., Musch, T., Stopper, H., Lyko, F., 2006. Functional diversity of DNA methyltransferase inhibitors in human cancer cell lines. Cancer Res. 66 (5), 2794−2800.

Struhl, K., Segal, E., 2013. Determinants of nucleosome positioning. Nat. Struct. Mol. Biol. 20 (3), 267−273.

Sun, X., Hung, K., Wu, L., Sidransky, D., Guo, B., 2002. Detection of tumor mutations in the presence of excess amounts of normal DNA. Nat. Biotechnol. 20 (2), 186−189.

Sun, Z., Chai, H.S., Wu, Y., White, W.M., Donkena, K.V., Klein, C.J., Garovic, V.D., Therneau, T.M., Kocher, J.P., 2011. Batch effect correction for genome-wide methylation data with Illumina Infinium platform. BMC Med. Genomics 4 (1), 84.

Sun, Z., Cunningham, J., Slager, S., Kocher, J.P., 2015. Base resolution methylome profiling: considerations in platform selection, data preprocessing and analysis. Epigenomics 7 (5), 813−828.

Suzuki, H., Gabrielson, E., Chen, W., Anbazhagan, R., van Engeland, M., Weijenberg, M.P., Herman, J.G., Baylin, S.B., 2002. A genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in human colorectal cancer. Nat. Genet. 31 (2), 141–149.

Tabatabai, G., Stupp, R., van den Bent, M.J., Hegi, M.E., Tonn, J.C., Wick, W., Weller, M., 2010. Molecular diagnostics of gliomas: the clinical perspective. Acta Neuropathol. 120 (5), 585–592.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., Rao, A., 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324 (5929), 930–935.

Taiwo, O., Wilson, G.A., Morris, T., Seisenberger, S., Reik, W., Pearce, D., Beck, S., Butcher, L.M., 2012. Methylome analysis using MeDIP-seq with low DNA concentrations. Nat. Protoc. 7 (4), 617–636.

Tak, Y.G., Farnham, P.J., 2015. Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. Epigenetics Chromatin 8, 57.

Taylor, K.H., Kramer, R.S., Davis, J.W., Guo, J., Duff, D.J., Xu, D., Caldwell, C.W., Shi, H., 2007. Ultradeep bisulfite sequencing analysis of DNA methylation patterns in multiple gene promoters by 454 sequencing. Cancer Res. 67 (18), 8511–8518.

Thakore, P.I., D'Ippolito, A.M., Song, L., Safi, A., Shivakumar, N.K., Kabadi, A.M., Reddy, T.E., Crawford, G.E., Gersbach, C.A., 2015. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. Nat. Methods 12.

Thomassin, H., Kress, C., Grange, T., 2004. MethylQuant: a sensitive method for quantifying methylation of specific cytosines within the genome. Nucleic Acids Res. 32 (21), e168.

Tobi, E.W., Goeman, J.J., Monajemi, R., Gu, H., Putter, H., Zhang, Y., Slieker, R.C., Stok, A.P., Thijssen, P.E., Muller, F., van Zwet, E.W., Bock, C., Meissner, A., Lumey, L.H., Slagboom, P.E., Heijmans, B.T., 2014. DNA methylation signatures link prenatal famine exposure to growth and metabolism. Nat. Commun. 5, 5592.

Tost, J., 2008. Epigenetics. Horizon Scientific Press, Norwich, UK.

Tost, J., 2016. Engineering of the epigenome: synthetic biology to define functional causality and develop innovative therapies. Epigenomics 8 (2), 153–156.

Tost, J., Dunker, J., Gut, I.G., 2003a. Analysis and quantification of multiple methylation variable positions in CpG islands by Pyrosequencing. Biotechniques 35 (1), 152–156.

Tost, J., El abdalaoui, H., Gut, I.G., 2006. Serial pyrosequencing for quantitative DNA methylation analysis. Biotechniques 40 (6), 721–722, 724, 726.

Tost, J., Gut, I.G., 2002. Genotyping single nucleotide polymorphisms by mass spectrometry. Mass Spectrom. Rev. 21 (6), 388–418.

Tost, J., Gut, I.G., 2006. DNA analysis by mass spectrometry-past, present and future. J. Mass Spectrom. 41 (8), 981–995.

Tost, J., Gut, I.G., 2007. DNA methylation analysis by pyrosequencing. Nat. Protoc. 2 (9), 2265–2275.

Toyota, M., Ho, C., Ahuja, N., Jair, K.W., Li, Q., Ohe-Toyota, M., Baylin, S.B., Issa, J.P., 1999. Identification of differentially methylated sequences in colorectal cancer by methylated CpG island amplification. Cancer Res. 59 (10), 2307–2312.

Tufarelli, C., Stanley, J.A., Garrick, D., Sharpe, J.A., Ayyub, H., Wood, W.G., Higgs, D.R., 2003. Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. Nat. Genet. 34 (2), 157–165.

Uhlmann, K., Brinckmann, A., Toliat, M.R., Ritter, H., Nürnberg, P., 2002. Evaluation of a potential epigenetic biomarker by quantitative methyl-single nucleotide polymorphism analysis. Electrophoresis 23 (24), 4072–4079.

Urich, M.A., Nery, J.R., Lister, R., Schmitz, R.J., Ecker, J.R., 2015. MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. Nat. Protoc. 10 (3), 475–483.

Ushijima, T., 2005. Detection and interpretation of altered methylation patterns in cancer cells. Nat. Rev. Cancer 5 (3), 223–231.

Ushijima, T., Morimura, K., Hosoya, Y., Okonogi, H., Tatematsu, M., Sugimura, T., Nagao, M., 1997. Establishment of methylation-sensitive-representational difference analysis and isolation of hypo- and hypermethylated genomic fragments in mouse liver tumors. Proc. Natl. Acad. Sci. U.S.A. 94 (6), 2284–2289.

Vaca-Paniagua, F., Oliver, J., Nogueira da Costa, A., Merle, P., McKay, J., Herceg, Z., Holmila, R., 2015. Targeted deep DNA methylation analysis of circulating cell-free DNA in plasma using massively parallel semiconductor sequencing. Epigenomics 7 (3), 353–362.

van den Bent, M.J., Dubbink, H.J., Sanson, M., van der Lee-Haarloo, C.R., Hegi, M., Jeuken, J.W.M., Ibdaih, A., Brandes, A.A., Taphoorn, M.J.B., Frenay, M., Lacombe, D., Gorlia, T., Dinjens, W.N.M., Kros, J.M., 2009. MGMT promoter methylation is prognostic but not predictive for outcome to adjuvant PCV chemotherapy in anaplastic oligodendroglial tumors: a report from EORTC Brain Tumor Group Study 26951. J. Clin. Oncol. 27 (35), 5881–5886.

Varley, K.E., Mitra, R.D., 2010. Bisulfite Patch PCR enables multiplexed sequencing of promoter methylation across cancer samples. Genome Res. 20 (9), 1279–1287.

Vicente, R., Noel, D., Pers, Y.M., Apparailly, F., Jorgensen, C., 2015. Deregulation and therapeutic potential of microRNAs in arthritic diseases. Nat. Rev. Rheumatol. 12 (4), 211–220.

Voelter-Mahlknecht, S., 2016. Epigenetic associations in relation to cardiovascular prevention and therapeutics. Clin. Epigenet. 8, 4.

von Deimling, A., Korshunov, A., Hartmann, C., 2011. The next generation of glioma biomarkers: MGMT methylation, BRAF fusions and IDH1 mutations. Brain Pathol. (Zurich, Switzerland) 21 (1), 74–87.

Waddington, C.H., 1942. The epigenotype. Endeavour 1, 18–20.

Wang, Q., Gu, L., Adey, A., Radlwimmer, B., Wang, W., Hovestadt, V., Bahr, M., Wolf, S., Shendure, J., Eils, R., Plass, C., Weichenhan, D., 2013. Tagmentation-based whole-genome bisulfite sequencing. Nat. Protoc. 8 (10), 2022–2032.

Wang, R.Y., Gehrke, C.W., Ehrlich, M., 1980. Comparison of bisulfite modification of 5-methyldeoxycytidine and deoxycytidine residues. Nucleic Acids Res. 8 (20), 4777–4790.

Wang, K., Li, X., Dong, S., Liang, J., Mao, F., Zeng, C., Wu, H., Cai, W., Sun, Z.S., 2015. Q-RRBS: a quantitative reduced representation bisulfite sequencing method for single-cell methylome analysis. Epigenetics 10 (9), 775–783.

Warnecke, P.M., Stirzaker, C., Melki, J.R., Millar, D.S., Paul, C.L., Clark, S.J., 1997. Detection and measurement of PCR bias in

quantitative methylation analysis of bisulphite-treated DNA. Nucleic Acids Res. 25 (21), 4422–4426.

Warnecke, P.M., Stirzaker, C., Song, J., Grunau, C., Melki, J.R., Clark, S.J., 2002. Identification and resolution of artifacts in bisulfite sequencing. Methods 27 (2), 101–107.

Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., Schubeler, D., 2005. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat. Genet. 37 (8), 853–862.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M., Schubeler, D., 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat. Genet. 39 (4), 457–466.

Wei, S.H., Balch, C., Paik, H.H., Kim, Y.-S., Baldwin, R.L., Liyanarachchi, S., Li, L., Wang, Z., Wan, J.C., Davuluri, R.V., Karlan, B.Y., Gifford, G., Brown, R., Kim, S., Huang, T.H.M., Nephew, K.P., 2006. Prognostic DNA methylation biomarkers in ovarian cancer. Clin. Cancer Res. 12 (9), 2788–2794.

Wei, S.H., Chen, C.-M., Strathdee, G., Harnsomburana, J., Shyu, C.-R., Rahmatpanah, F., Shi, H., Ng, S.-W., Yan, P.S., Nephew, K.P., Brown, R., Huang, T.H.-M., 2002. Methylation microarray analysis of late-stage ovarian carcinomas distinguishes progression-free survival in patients and identifies candidate epigenetic markers. Clin. Cancer Res. 8 (7), 2246–2252.

Weisenberger, D.J., Siegmund, K.D., Campan, M., Young, J., Long, T.I., Faasse, M.A., Kang, G.H., Widschwendter, M., Weener, D., Buchanan, D., Koh, H., Simms, L., Barker, M., Leggett, B., Levine, J., Kim, M., French, A.J., Thibodeau, S.N., Jass, J., Haile, R., Laird, P.W., 2006. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. Nat. Genet. 38 (7), 787–793.

White, H.E., Hall, V.J., Cross, N.C., 2007. Methylation-sensitive high-resolution melting-curve analysis of the SNRPN gene as a diagnostic screen for Prader-Willi and Angelman syndromes. Clin. Chem. 53 (11), 1960–1962.

Wick, W., Hartmann, C., Engel, C., Stoffels, M., Felsberg, J., Stockhammer, F., Sabel, M.C., Koeppen, S., Ketter, R., Meyermann, R., Rapp, M., Meisner, C., Kortmann, R.D., Pietsch, T., Wiestler, O.D., Ernemann, U., Bamberg, M., Reifenberger, G., von Deimling, A., Weller, M., 2009. NOA-04 randomized phase III trial of sequential radiochemotherapy of anaplastic glioma with procarbazine, lomustine, and vincristine or temozolomide. J. Clin. Oncol. 27 (35), 5874–5880.

Wiebers, J.L., 1976. Detection and identification of minor nucleotides in intact deoxyribonucleic acids by mass spectrometry. Nucleic Acids Res. 3 (11), 2959–2970.

Wielscher, M., Vierlinger, K., Kegler, U., Ziesche, R., Gsur, A., Weinhausel, A., 2015. Diagnostic performance of plasma DNA methylation profiles in lung cancer, pulmonary fibrosis and COPD. EBioMedicine 2 (8), 927–934.

Wojdacz, T.K., Dobrovic, A., 2007. Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation. Nucleic Acids Res. 35 (6), e41.

Wojdacz, T.K., Dobrovic, A., Algar, E.M., 2008. Rapid detection of methylation change at H19 in human imprinting disorders using methylation-sensitive high-resolution melting. Hum. Mutat. 29 (10), 1255–1260.

Worm, J., Aggerholm, A., Guldberg, P., 2001. In-tube DNA methylation profiling by fluorescence melting curve analysis. Clin. Chem. 47 (7), 1183–1189.

Xiao, Y., Yu, F., Pang, L., Zhao, H., Liu, L., Zhang, G., Liu, T., Zhang, H., Fan, H., Zhang, Y., Pang, B., Li, X., 2015. MeSiC: a model-based method for estimating 5 mC levels at single-CpG resolution from MeDIP-seq. Sci. Rep. 5, 14699.

Xiong, Z., Laird, P.W., 1997. COBRA: a sensitive and quantitative DNA methylation assay. Nucleic Acids Res. 25 (12), 2532–2534.

Xu, X., Zhang, Y., Williams, J., Antoniou, E., McCombie, W.R., Wu, S., Zhu, W., Davidson, N.O., Denoya, P., Li, E., 2013. Parallel comparison of illumina RNA-Seq and affymetrix microarray platforms on transcriptomic profiles generated from 5-aza-deoxy-cytidine treated HT-29 colon cancer cells and simulated datasets. BMC Bioinforma. 14 (Suppl. 9), S1.

Yamada, Y., Watanabe, H., Miura, F., Soejima, H., Uchiyama, M., Iwasaka, T., Mukai, T., Sakaki, Y., Ito, T., 2004. A comprehensive analysis of allelic methylation status of CpG islands on human chromosome 21q. Genome Res. 14 (2), 247–266.

Yang, A.S., Estecio, M.R., Doshi, K., Kondo, Y., Tajara, E.H., Issa, J.P., 2004. A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. Nucleic Acids Res. 32 (3), e38.

Yi, J.M., Dhir, M., Van Neste, L., Downing, S.R., Jeschke, J., Glöckner, S.C., de Freitas Calmon, M., Hooker, C.M., Funes, J.M., Boshoff, C., Smits, K.M., van Engeland, M., Weijenberg, M.P., Iacobuzio-Donahue, C.A., Herman, J.G., Schuebel, K.E., Baylin, S.B., Ahuja, N., 2011. Genomic and epigenomic integration identifies a prognostic signature in colon cancer. Clin. Cancer Res. 17 (6), 1535–1545.

Yoder, J.A., Walsh, C.P., Bestor, T.H., 1997. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. 13 (8), 335–340.

Yoo, C.B., Jones, P.A., 2006. Epigenetic therapy of cancer: past, present and future. Nat. Rev. Drug Discov. 5 (1), 37–50.

Yu, L., Chibnik, L.B., Srivastava, G.P., Pochet, N., Yang, J., Xu, J., Kozubek, J., Obholzer, N., Leurgans, S.E., Schneider, J.A., Meissner, A., De Jager, P.L., Bennett, D.A., 2015. Association of brain DNA methylation in SORL1, ABCA7, HLA-DRB5, SLC24A4, and BIN1 with pathological diagnosis of Alzheimer disease. JAMA Neurol. 72 (1), 15–24.

Yu, S.L., Chen, H.Y., Chang, G.C., Chen, C.Y., Chen, H.W., Singh, S., Cheng, C.L., Yu, C.J., Lee, Y.C., Chen, H.S., Su, T.J., Chiang, C.C., Li, H.N., Hong, Q.S., Su, H.Y., Chen, C.C., Chen, W.J., Liu, C.C., Chan, W.K., Chen, W.J., Li, K.C., Chen, J.J., Yang, P.C., 2008. MicroRNA signature predicts survival and relapse in lung cancer. Cancer Cell 13 (1), 48–57.

Yuan, E., Haghighi, F., White, S., Costa, R., McMinn, J., Chun, K., Minden, M., Tycko, B., 2006. A single nucleotide polymorphism chip-based method for combined genetic and epigenetic profiling: validation in decitabine therapy and tumor/normal comparisons. Cancer Res. 66 (7), 3443–3451.

Yuan, Z., Claros, C.S., Suzuki, M., Maggi, E.C., Kaner, J.D., Kinstlinger, N., Gorecka, J., Quinn, T.J., Geha, R., Corn, A., Pastoriza, J., Jing, Q., Adem, A., Wu, H., Alemu, G., Du, Y.C., Zheng, D., Greally, J.M., Libutti, S.K., 2016. Loss of MEN1 activates DNMT1 implicating DNA hypermethylation as a driver of MEN1 tumorigenesis. Oncotarget 7.

Zeilinger, S., Kuhnel, B., Klopp, N., Baurecht, H., Kleinschmidt, A., Gieger, C., Weidinger, S., Lattka, E., Adamski, J., Peters, A., Strauch, K., Waldenberger, M., Illig, T., 2013. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PLoS One 8 (5), e63812.

Zeschnigk, M., Bohringer, S., Price, E.A., Onadim, Z., Masshofer, L., Lohmann, D.R., 2004. A novel real-time PCR assay for quantitative analysis of methylated alleles (QAMA): analysis of the retinoblastoma locus. Nucleic Acids Res. 32 (16), e125.

Zhang, Y., Yang, R., Burwinkel, B., Breitling, L.P., Brenner, H., 2014. F2RL3 methylation as a biomarker of current and lifetime smoking exposures. Environ. Health Perspect. 122 (2), 131−137.

Chapter 8

Perspectives for Future DNA Sequencing Techniques and Applications

W.J. Ansorge[1], T. Katsila[2] and G.P. Patrinos[2]
[1]*Ecole Polytechnique Federal Lausanne, EPFL, Lausanne, Switzerland;* [2]*University of Patras School of Health Sciences, Patras, Greece*

8.1 INTRODUCTION

In the past few years (Ansorge, 2009, 2010), DNA sequencing techniques and their applications in the diagnostics field have made significant progress. The innovations are reflected particularly in higher throughput and lower cost per base, e.g., on Illumina and Ion Torrent systems, or the single-molecule real-time sequencing approaches of Pacific Bioscience and Oxford Nanopore. The human genome sequence has complemented the understanding of biology, human diversity, and disease. Advances in DNA sequencing technologies over the past decade or so have made possible the transition from the first human genome sequences to personal genomes and genomic medicine. Next-generation sequencing (NGS) uses massively parallel sequencing to generate up to several gigabases of sequence information per day, making possible new applications that were previously difficult to perform. The technique has spread to almost all fields concerning DNA: food control, screening of the purity of vaccines and other medications (detection of traces of the active viral components and nonactive viral particles), evolution of organisms, analysis of fetus DNA from the mother's blood, the forensic field, and others.

Several reviews have described the historical development of the field (Ansorge, 2009, 2010). Here, we focus on the status of the technology in the second half of 2015, both of novel commercially available platforms and of systems in development. Techniques are discussed that have the potential for future DNA sequencing techniques, as well as some of the unsolved challenges for the platforms, the price to performance ratio, the complexity of sample preparation, and the detection of structural variations in genome. In the past few years rapidly developing technologies and methods have permitted analysis of the genome and transcriptome of a single cell. The first observations suggest that both genomic and transcriptomic heterogeneities within an organism are more common than expected during normal development and disease.

The diagnostics field has already profited from new generations of DNA sequencing technology, sample preparation, single-cell isolation techniques, DNA amplification schemes, and computer data analysis with improving software tools. The applications resulted in better understanding of causes of disease and elucidation of reasons why a drug efficiently treats a disease in some persons whereas the treatment is unsuccessful in others with the same disease. Recent public support for precision medicine, novel therapeutic approaches, and efforts for improvements in health care will be motivations for further innovations and developments in the field.

The aims for the technology will be to lower the cost of the equipment and biochemicals involved, simultaneously increasing the reproducibility, reliability, and simplicity of the techniques and protocols in operation. The goal to be achieved, formulated in Ansorge (2009, 2010) is still valid despite the great progress in DNA sequencing techniques since then. For genomic sequencing, and for analysis of the ever more important structural genetic variations in genomes [e.g., copy number variations (CNVs), for chromosomal translocations, inversions, large deletions, insertions, and duplications], it would be a great advantage if sequence read length on the original single DNA molecule could be increased to tens of thousands of bases and more in a relatively short time. Ideally the goal would be the sequence determination of a whole chromosome, from a single original DNA molecule, from a single cell.

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00008-0
Copyright © 2017 Elsevier Ltd. All rights reserved.

8.2 COMMERCIALLY AVAILABLE ANALYSIS PLATFORMS

So far (Ansorge, 2009, 2010), several new companies have entered the NGS field (e.g., Oxford Nanopore, Genia, GnuBIO); their devices are described in the following sections.

Some companies limited the distribution of their systems or are no longer active in the market. The first was Helicos, and although it was a pioneer in the field, its revolutionary single molecule detection technique used in the device was not yet mature enough and production stopped. The technique was recently redesigned, and a prototype of the new device called GenoCare was developed and presented in October 2015 by Direct Genomics (see Section 3.3).The SOLiD system platform from Life Sciences did not find decisive favor with users, possibly because of the complex biochemical sample preparation, high cost, and relatively short reading length, around 100 bases. Roche stopped production and distribution of the pyrosequencing bioluminescence detection Gene Sequencer because of its higher sequencing cost, which was not competitive. It was originally developed by the 454 company as the first system from the Next Generation devices on the market, and for several years it was applied successfully in many projects. After the attempt to take over Illumina, Roche initiated collaborations with sequencing platforms of Pacific Biosystems and Genia companies, discussed subsequently.

8.2.1 Illumina Platform

The company (www.illumina.com) successfully pushed its technology, originally developed by Solexa (Ansorge, 2009, 2010), and continuously introduced novel devices for very high throughput as well as smaller and more modestly priced versions. At the same time, it improved the reading length from the originally relatively short 50 bases to about 300 bases, improved the biochemistry of the sample preparation several times, and reduced the sequencing cost down to $0.001 per 1000 bases, which would not have been expected a few years earlier.

The sequencing devices offer starts with the MiSeq series, with speed and simplicity suitable for small genomes, amplicons, and targeted gene panel sequencing, and it is also foreseen for applications in clinical research. The system has one flow cell; the output ranges from 0.3 to 15 Gb, the number of reads per flow cell is 25 million, the run time is 5−55 h, and the maximum read length is 2× 300 base pairs (bp).

The HiSeq series in the middle range is used for sequencing of genomes, exomes, and transcriptomes. As an example, the HiSeq 2500 device in the rapid mode has one or two flow cells per run; the output range is 10−300 Gb, the run time is 7−60 h; it results in 300 million reads per flow cell, and the maximum read length is 2× 250 bp.

The HiSeq X series with very high throughput is used on the production and population scale for whole-genome sequencing. As an example, the HiSeq 4000 device has one or two flow cells per run; the output range is 125−1500 Gb, the run time is 1−3.5 days, it has 2.5 milliards (in the US, 2.5 billion) reads per flow cell, and the maximum read length is 2× 150 bp.

The data flow resulting from the large sequencing throughput and its analysis requires great effort on the part of bioinformatics. One cause may be the still relatively short reads obtainable with the technique.

There are many Illumina systems purchased by laboratories worldwide. It is possible that the global sequencing capacity is sufficiently large for current needs, so the next wave of acquisitions may not grow as fast as the previous one, taking into account the competing systems being introduced into operation, as described subsequently.

The schema of operation, the numerous applications in analysis of gene expression and genomes, relevant publications, collaborations, and operation protocols involving the Illumina systems are listed at the website of the company (www.illumina.com). The opinion of users can be found on the Genome website (www.genomeweb.com).

8.2.2 Ion Torrent: Thermo Fisher Scientific

Ion Torrent technology has been a continuation of the original pyrosequencing technique, as explained, e.g., in Ansorge (2009, 2010). In its original design, the 454 company detected the pyrophosphate group released during the incorporation of a nucleotide in the complementary strand during the DNA polymerization process. The current Ion Torrent system detects H^+ ions, another product released during DNA polymerization. The basic principle, the detection of H^+ ions, has been licensed from DNA Electronics in London. The technology combines semiconductor sequencing technology with biochemistry, enabling the direct translation of chemical information into digital sequence data. This eliminates the need for expensive optics, lasers, and complex sequencing chemistries with fluorescently labeled nucleotides. The result is a sequencing system that is more affordable and cost-effective, faster, scalable, and relatively simple to operate.

The Ion PGM System (depending on the chip used and the application goals) can run up to 5.5 million reads, with an output reaching 2 Gb, a reading length up to 400 bases, and a run time between 2 and 7 h.

The Ion Proton System has been designed for high-throughput sequencing of exomes, transcriptomes, and genomes. With the Ion PI Chip, the throughput is up to 10 Gb, the number of reads is up to 80 million, the read length is 200 bases, and the run time is 2−4 h. It has

applications in the sequencing of genes, human genomes, de novo sequencing, chromatin immunoprecipitation technique, whole transcriptomes, exomes, methylation analysis, gene expression by sequencing, small genomes, and small RNAs.

The next-generation sequencing devices Ion S5 and Ion S5 XL are being introduced for a simple and rapid workflow for panels, microbes, exomes, and transcriptomes, and are applicable both for high and low weekly throughput. Output with the Ion 540 Chip is up to 15 Gb (about half with reads of 400 bases); the number of reads is up to 80 millions, and the run time is 2.5−5 h (for reads of 200 or 400 bases). The claim is that only 10 ng of low-quality DNA or RNA is needed to generate mutational or gene expression profiles. Target selection is fast and simple using Ion AmpliSeq panels, also requiring only as little as 10 ng of input material. Ion Torrent systems have simplified NGS data analysis with end-to-end bioinformatics solutions, including the Ion Reporter Software and server. Operating software and Plugins are available for analysis of genome sequencing, targeted sequencing, variant analysis and annotation, microbial sequencing, transfer to third-party bioinformatics packages, and more.

Because of their affordable price and simplicity of operation, Ion Torrent systems are used worldwide in many projects. The schema of operation and list of publications and collaborations using this system are obtainable from the company Web page (www.thermofisher.com). Opinions of users can be found on the Genome website (www.genomeweb.com).

8.2.3 Pacific Biosystems

Pacific Biosystems has been developing single-molecule real-time (SMRT) technology for many years. As the technology improved continuously over the years it successfully attracted an increasing number of users. The PacBio Sequencing System may provide very high amount of genetic information through exceptionally long sequencing reads (average of >10,000 bp) and high-consensus accuracy.

The basic principle of the technology, explained in www.pacbio.com, is detection of the fluorescently labeled nucleotide as it is being incorporated in the complementary strand of the single molecule by the DNA polymerase. In the moment of incorporation the fluorescent label specific for the base is detected, while the polymerase simultaneously cuts off the label from the nucleotide. The process is repeated for the next labeled nucleotide, and from the order of the four different labels detected is successively determined the base sequence. The process takes place in zero-mode waveguide (ZMW) chambers. Within the ZMWs, a single molecule of polymerase is immobilized and fluorescently labeled nucleotide is added so that DNA sequencing can be observed optically and recorded in real time.

In the first **PacBio RS II** system the SMRT cell contains 150,000 ZMWs, with a throughput of 1 Gb per SMRT cell. It has high-performance optics, automated liquid handling, SMRT cells and reagents, instrument control software, a run time from 30 min to 6 h per SMRT cell, and a run size from 1 to 16 SMRT cells per run. With this established technology, there are over 850 scientific publications in a variety of research areas such as human biomedical research, plant and animal genomics, and microbiology. The PacBio RS II has been applied to whole-genome sequencing of small genomes, targeted sequencing, complex population analysis, RNA sequencing of targeted transcripts, and microbial epigenetics.

In autumn 2015, Pacific Biosciences launched a new **Sequel** single-molecule sequencing system that offers higher throughput and costs at about half the price of the original RS II device. PacBio developed the Sequel as part of its collaboration with Roche to develop a clinical-grade sequencing system for diagnostic purposes. The device for the Roche sequencing instrument is expected to be launched in 2016 for clinical research, and later as an in vitro diagnostics instrument. The Sequel system features redesigned SMRT cells, each of which contains 1 million ZMWs, compared with the 150,000 ZMWs in the RS II device. The increased number of ZMWs per SMRT cell enables an approximately sevenfold improvement in throughput, so the total will range between 5 and 10 Gb and the system has an average read length between 8 and 12 kb. Like the RS II, the Sequel can run between 1 and 16 SMRT cells per run. Sample preparation will be the same on both instruments, taking around 6 h to prepare a library. The chemistry is the same as in the RS II; therefore the consensus accuracy will be the same. The Sequel is equipped with improved optics with a computer system that processes more data. It supports all applications that were used in the RS II device, including transcript and base modification analysis.

Although the consumables cost per SMRT cell will increase for the Sequel (to around $700 from about $250 for the RS II), scientists will obtain sevenfold improvement in overall throughput. Assuming 7 Gb of data per SMRT cell, PacBio estimates that sequencing a human genome at 10× coverage [which should be deep enough to do structural variant (SV) analysis] would cost around $3000. A 30× human genome would cost around $10,000 and a 30× human genome would cost $15,000. (For comparison, in the previous RS II system the SMRT cell contains 150,000 ZMWs, with a throughput of 1 Gb per SMRT cell, a 30× human genome would cost around $22,500 whereas a 50× genome would cost $37,500.)

The price to sequence a human genome on the Sequel will be still higher compared with sequencing on the

Illumina, but the cost for computer and analysis time may be lower owing to longer reads and may result in easier assembly. With its long reads, it will be used in projects for rapid and cost-effective generation of whole-genome de novo assemblies, particularly of plant genomes, or in projects analyzing hard-to-sequence regions of the human genome. The diagnostic instrument (in operation in 2016, with Roche collaboration) is planned for the clinical market.

Schema of operation, the numerous applications in analysis of genes and genomes, relevant publications, collaborations, and operation protocols involving the PacBio system are listed on the website of the company (www.pacbio.com). Opinions of users and the latest news can be found on the Genome website (www.genomeweb.com).

8.2.4 Complete Genomics and BGI

Complete Genomics, founded in 2006, plans to become one of the leaders in whole-human genome sequencing. Its technology (Drmanac et al., 2010) is based on shearing human DNA into pieces about 400 bases long, then using rolling circle amplification to produce so-called nanoballs, which are then spotted on a patterned substrate. The patterned substrate with nanoballs is subsequently exposed to a repeated proprietary hybridization technique by ligation with fluorescent labeled probes. The base sequence is determined from the fluorescence signals resulting from the repeated hybridization on the nanoballs. The company has developed for its technique the sequencing instruments, chemistry, and software. Complete genomics has sequenced more than 20,000 whole human genomes over 5 years. In a study published in *Nature* (Brock et al., 2012), Complete Genomics demonstrated its long fragment read technology, which enables whole-genome sequencing and haplotyping from 10 to 20 cells with an error rate of 1 in 10 million bases, which it hopes would enable truly clinical-grade genomes.

In March 2013, Complete Genomics was acquired by BGI-Shenzhen, the world's largest genomics services company, with headquarters in Shenzhen, China. It provides comprehensive sequencing and bioinformatics services for commercial science, medical, agricultural, and environmental applications, and collaborates globally in many projects with leading research institutions. Before acquisition by BGI, Complete Genomics was focused on selling whole-genome sequencing services to over 150 research customers. It now focuses on building a new generation of high-throughput sequencing technology for clinical and consumer applications. In 2015, Complete Genomics produced and introduced the **Revolocity** system, the end-to-end genomics system for large-scale genomes. The Revolocity system will use a 300-base insert and mate pair sequencing of 28 bases from each end of the insert. The older technology with the four-adaptor construction had mate pair reads of 35 bases (Drmanac et al., 2010). The remainder of the sequencing technology is the same. DNA nanoballs are formed from the circular pieces of DNA and deposited onto patterned silicon arrays and sequenced by ligation.

Revolocity enables 96% of the genome to be covered with an error rate of about 10^{-6}. The system will initially be applied to exome and whole-genome sequencing but the company plans to develop other applications such as RNA sequencing.

In October 2015, BGI launched a new desktop sequencing platform, **BGISEQ-500**, based on the technology of Complete Genomics. It uses a variation on the "DNA nanoballs" sequencing method, with new engineering features. Whereas Complete Genomics' instruments are applied to sequencing whole human genomes and exomes, the new instrument is more flexible and is planned for RNA sequencing, including in single cells, panels for noninvasive prenatal sequencing, genotyping panels, and other applications. It is a benchtop machine with high-throughput, about 200 Gb per run. The BGISEQ-500 will also be capable of rapid runs producing as little as 8 Gb of data, an advantage for clinical assays.

BGI and its customers have generated over 250 publications in top journals. Schema of operation, the numerous applications in analysis of genes and genomes, relevant publications, collaborations, and operation protocols involving Complete Genomics and BGI systems are listed at the website of the companies (www.completegenomics.com). Opinions of users and the latest news can be found on the Genome website (www.genomeweb.com).

8.2.5 Oxford Nanopore Platform

The principle of operation of the nanopore sequence technique is the analysis of the DNA strand directly as the molecule is drawn through a tiny pore suspended in a membrane. Changes in electrical current, or tunneling currents, are used to read off the chain of bases. Great hope and attention are devoted to emerging nanopore technology because it would allow to single DNA molecule to be sequenced in real-time, requiring no amplification step, combined with less expensive hardware and simplicity of operation.

The promise is great, and progress is finally taking place slowly. When Oxford Nanopore presented the **MinIon device** in 2012, it was met with slightly pessimistic expectations. In the meantime, many laboratories worldwide are part of the MinIon Analysis and Reference Consortium (MARC), a group formed by a number of participants in Oxford Nanopore's MinIon early-access program that plans to conduct a series of projects based on the technology. They all wish the technology to be a success and are assessing how reproducible the device's performance is across laboratories, and are developing standard protocols and reference data.

The consortium made the raw and aligned nanopore reads of the first test results available (corresponding to the state of the MinIon technology in April 2015; see www.genomeweb.com, October 15, 2015). The European Nucleotide Archive and the European Bioinformatics Institute coordinated the data distribution and analysis. The goals of the study were to assess the yield, accuracy, and reproducibility of the MinIon by performing replicate experiments at several sites and to find out which technical factors determine high performance. For the study, each laboratory generated MinIon sequence data from the same substrain of *Escherichia coli* using the same protocols.

Overall, using R7.3 flow cells and SQK-MAP005 chemistry, a typical experiment resulted in about 20,000 two-dimensional reads, with a median length of 6.5 kilobases, generating about 115 megabases of data. When they used an 8-kilobase shearing protocol, almost 5% of two-dimensional (2D) reads were at least 10,000 bases long, some of them more than 50,000 bases. The total error for individual 2D base calls was 12%, consisting of 3% miscalls; a further 4% were insertions and 5% were deletions. A single run yielded enough 2D bases to cover the *E. coli* genome 25-fold.

The performance of the MinIon device itself was consistent in the tests and no experiment failed owing to problems with the device. The researchers also observed no guanine-cytosine bias, although they noted this may be hard to detect with an *E. coli* genome. One reason the quantity and quality of data varied between runs is that many steps in the standard protocol are sensitive to the quality of materials and reagents being used. The results of the MARC report suggested how to improve the performance of the MinIon, including clearer protocol steps, methods for longer library molecules, and improved operation run software scripts.

Overall, MinIon nanopore sequencing has several advantages over short-read sequencing technologies, e.g., its long reads enable sequencing through repetitive regions, its generation in speed results in minutes rather than hours, it has a low capital cost of $1000 per device, and it has a small size. Its main limitations are the currently high mismatch and indel error rate and the limited yield per run, ranging from megabases to gigabases. However both will continue to improve through better bioinformatic tools and new hardware and chemistry.

Oxford Nanopore released the new **MinIon Mk1** device and reagents, and the company is constructing systems with larger throughput. Still, the priority is to concentrate on reproducibility and standardization of the performance.

8.2.5.1 Detection of Structural Variants With Oxford Nanopore MinIon Device

Nanopore sequencing can detect SVs in a mixture of polymerase chain reaction amplicons. In a proof-of-principle, the scientists, led by Jim Eshleman and Winston Timp from Johns Hopkins University, showed (GenomeWeb.com, October 5, 2015) that they could detect a number of well-characterized structural variants, including large deletions, inversions, and translocations, that affect two tumor suppressor genes in pancreatic cancer cell lines, using amplicon sequence data from Oxford Nanopore's MinIon. It shows the ability of nanopore sequencing to detect SVs correctly and reliably with only hundreds of reads instead of millions.

In detecting SVs, Oxford Nanopore is still behind other technologies that offer long-range genome information, such as Pacific Biosciences, BioNano Genomics (see subsequent discussion), Illumina synthetic long reads, and OpGen optical mapping. However the MinIon has more potential for diagnostic and screening use. The cost of the Oxford Nanopore platform and its operation are low compared with other systems in use. This will push sequencing to spread from research laboratories to the clinics, where nanopore sequencing could become the tool for the low-level detection of cancer-associated SVs and their early detection. Clinical applications will require improved throughput (compared with amplicon sequencing) needed both for genome-wide detection of SVs and targeted sequencing of regions likely to contain an SV. Increase in throughput will be accomplished with advances in the current MinIon platform and with the larger instruments that Oxford Nanopore is developing.

An increasing number of publications is using the technique, with applications in analysis of clinical samples and small genomes. Schema of the principle of operation, relevant publications, collaborations, and operation protocols involving the Oxford Nanopore system are listed on the website of the company (www.oxfordnanopore.com). Opinions of users and the latest news can be found on the Genome website (www.genomeweb.com).

8.2.6 Other Projects With Protein Nanopores and Solid-State and Graphene Nanopores

Nanopore systems in operation (Oxford Nanopore and Genia) use the protein nanopore channels in a membrane. It has been demonstrated that using special protein channels can give the resolution of single DNA bases (Manrao et al., 2012; Cherf et al., 2012; Laszlo et al., 2014). Efforts are under way to develop nanopores with solid-state techniques used in the semiconductor industry, which could have some advantages in the mass manufacturing process of chips with a large number of nanopores, with the ambition of increasing the reproducibility of nanopore performance. Currently the thickness of solid-state membranes is too large, 10−20 nm, to allow the resolution of a single DNA

base, because with this thickness there may be 30–50 bases simultaneously inside the nanopore. A further difficulty with current techniques is the large speed with which the DNA molecule passes through the solid nanopore under the influence of an applied electrical field, which makes it impossible for the electronic detection system to achieve single base resolution.

To slow and control the translocation speed of the single-strand DNA through the nanopore, one possible way to use DNA polymerases, as proposed, e.g., in 2010 at EPFL in Lausanne (Ansorge W., note about patent proposal for DNA polymerase as a motor for controlled transport of DNA strand through a nanopore, in an EPFL internal email communication on August 4, 2010 to Prof. L. Forro and Prof. A. Radenovic, and emailed to Prof. Hagan Bayley on August 17, 2010) and in a publication in 2012 (Cherf et al., 2012).

In an effort to reduce nanopore thickness, several projects were described using graphene material for the nanopore (Merchant et al., 2010; Schneider et al., 2010; Garaj et al., 2010; Drndic, 2014). The thickness of the graphene nanopore is so small that just about one-third of a nucleotide is inside the pore during the translocation of the DNA molecule through the pore, presenting the theoretical possibility of scanning the DNA molecule with a single-base resolution. However, as reported in the reports on graphene pores mentioned previously, there are unsolved obstacles. Among them is the large speed of DNA translocation requiring the design of much faster electronics signal detection, or the development of techniques allowing the speed to be lowered, e.g., by some controlled interactions of the passing DNA strand and the walls of the pore. Also reported were wetting problems of the graphene pore possibly caused by the hydrophobic surface, perhaps requiring some surface passivation process for the pore.

Other thin materials were tested for the nanopore membranes, e.g., hafnium oxide and boron nitride (Drndic, 2014), and molybdenum disulfide (Liu et al., 2014).

Besides measuring changes in the ionic current as the DNA passes through the pore, there are efforts to measure the electric signals perpendicular to the DNA molecule (called tunneling currents DNA sequencing) as it traverses the pore or channel. The expectations are that each base will show a sufficiently different tunneling current, depending on its electronic structure, and the base sequence will be deduced from differences in the tunneling currents.

8.2.7 Genia: Roche

As the various platforms strive toward the $1000 genome, Genia (www.genia.com) is looking beyond toward the $100 genome. The technology has the potential to reduce the price of sequencing and increase speed, accuracy, and sensitivity by moving away from complex sample preparation and optical detection. The basis of Genia's technology is the biological nanopore, a protein pore embedded in a lipid bilayer membrane. The planar electronic sensor technology enables highly efficient accuracy of current readings, which is a common limitation in nanopore sequencing efforts. Genia's NanoTag sequencing approach, developed in collaboration with Columbia and Harvard Universities (Kumar et al., 2012), uses a DNA replication enzyme to sequence a template strand with single-base precision, as base-specific engineered tags cleaved by the enzyme are captured by the nanopore. As the cleaved tags travel through the pore, they attenuate the current flow across the membrane in a sequence-dependent manner (Fig. 8.1). Thus the tag, not the nucleotide, passes through the pore.

Electrochemically, each of the four tag types interacts with the nanopore recognition site differently, partially blocking the ion current by a characteristic amount that results in a tag-specific electronic signature (Fig. 8.2). DNA sequences are computed from the residual currents measured on the nanopore–DNA complex during passage of the cleaved tags through the pore. The high sensitivity of electronic detection circuitry underneath each sensor enables the sequence of single DNA molecules to be obtained with the Genia platform. In the future the nanotags on the four nucleotides can be optimized with the aim of achieving the highest resolution of the single bases (Kumar et al., 2012).

Many of today's sequencing platforms rely on specialized, expensive optical sensors. Genia technology senses

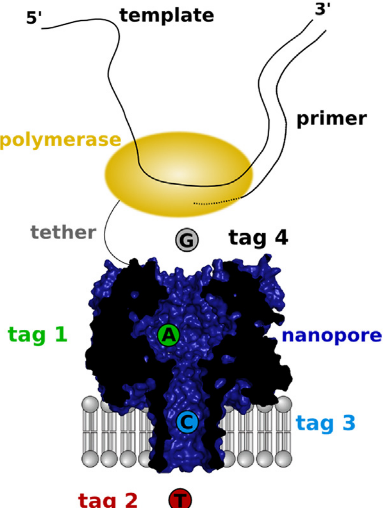


FIGURE 8.1 Schematic of a single molecule DNA sequencing by a nanopore with phosphate-tagged nucleotides. Genia's NanoTag sequencing approach identifies DNA sequences not by detecting the nucleotides themselves with the nanopore, but by measuring current changes caused by the passage of each of four different tags that are released from the incorporated nucleotide during the polymerase reaction. *From the Genia Web page, www.genia.com.*

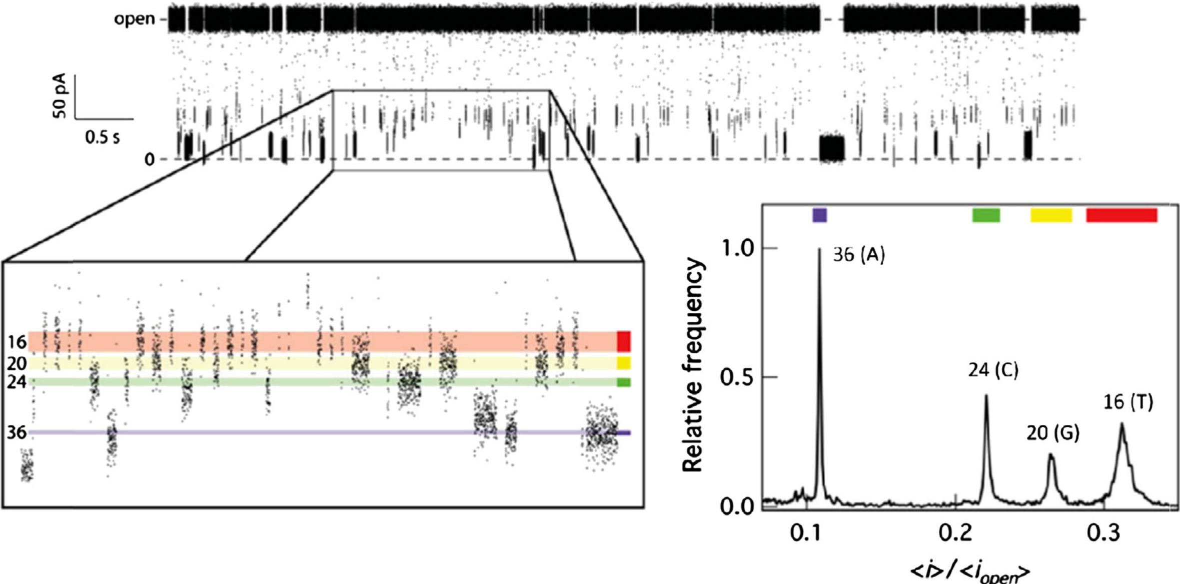


FIGURE 8.2 Example using unique current blockade signatures from engineered electronic tags to identify individual nucleotides. *From:* Scientific Reports *(Nature Publications group), "PEG-labeled nucleotides and nanopore detection for single molecule DNA sequencing by synthesis" (2012, 2, 684;* http://dx.doi.org/10.1038/srep00684*) and the Genia Web page* www.genia.com.

changes in electrical currents detected by standard semiconductors devices, without optical systems. Involving semiconductor technology with the DNA sequencing will reduce the cost of Genia's sequencer operation, making the $100 genome a possibility. The company was acquired by Roche.

8.2.8 BioNano Genomics

BioNano Genomics is not producing classical DNA sequencing devices, although it uses similar nanochannel technology and fluorescence detection. By providing a detailed physical Genome Map, however, it helps to finish sequencing and remove sequencing errors caused by repetitive regions. Genome Maps are a complement to sequencing and increase the quality of the results. NGS technologies are essential for nucleotide-level information, but some are limited in long-range resolution. The reason is that often the fragmented reads are too short compared with repetitive regions, and a longer base reading length would be needed to make a complete map and overcome repetitive regions within and between genes. To detect variations and architecture on a larger scale, scientists need a technology to map the genome directly with a resolution matching the variation in question, e.g., kilobases, not bp. Adding a genome map to sequencing data enables a better view of the whole genome, showing its features in context and functional relationships across kilobases to megabases. Sequencing can be completed to a higher standard in less time, with the help of a physical map to facilitate de novo sequence assembly and scaffolding.

The high-resolution **Irys** System from BioNano Genomics produces a detailed whole-genome map with highly precise anchor points against which to align data from NGS. Genome maps orient contigs and size gaps by bridging across repeats and other complex elements that break NGS assemblies. The system improves the long-range contiguity across the genome and in many cases helped to find errors in sequence assembly. Genome Maps represent a useful validation method for genome contigs and scaffolds in regions assembled from fragment reads.

SVs in chromosomal DNA account for a large amount of variability between human genomes, also influencing phenotypic variations. In clinical research, SVs are increasingly associated with several diseases such as Crohn disease, autism, schizophrenia, morbid obesity, and cancer. To study SV, researchers have traditionally been limited to techniques in which the genomic DNA is first sheared into small fragments, which limits the correct assembly of complex genomes. During assembly, repeats and structural relationships between regions are lost by fragmentation. In the Irys System excessive fragmentation is avoided by relaxing and straightening large DNA molecules in nanochannels so they can be imaged intact. Analysis of large whole molecules and generation of de novo genome maps allow the elements in the genome to be viewed directly and SVs to be analyzed. Other technologies that offer long-range genome information are Pacific Biosciences, Illumina synthetic long reads, OpGen optical mapping, and Oxford Nanopore.

In the Irys System, the DNA to be analyzed is labeled at specific sequence motifs for imaging and identification in

IrysChips. The labeling method in the system uses a nicking endonuclease to create a single-strand cut in the long DNA molecules, at a specific recognition site, wherever it occurs in the genome. Fluorescently labeled nucleotides are then incorporated by repairing the single-strand nicks with a polymerase. These labeling steps result in a uniquely identifiable sequence-specific pattern of labels to be used for de novo map assembly or for anchoring sequencing contigs. Other labeling techniques may be used in the system. The labeled DNA sample is pipetted onto the IrysChip in the flowcells; movement of DNA in the flowcell is controlled electrophoretically. The DNA stretches in solution to confine chromosomal-length DNA inside the nanochannels. The current is transiently turned off and the molecules in solution are stationary and stretched uniformly. At this time imaging is performed, capturing high-resolution, single-molecule images of the labeled DNA that contains sequence motifs along hundreds of kilobases (or even over a megabase) in a single contiguous molecule. Once imaged, the molecules are flushed and the process is repeated, allowing imaging of several gigabases of DNA per hour.

This technique reveals meaningful biological information despite the presence of repetitive elements that cannot be resolved by other technologies. It is also possible to detect epigenetic modifications across the genome. In contrast, in techniques in which the DNA molecules are sheared to shorter fragments, biological information may be disrupted or completely lost, e.g., the order and arrangement of functional regions.

Raw image data of labeled long DNA molecules are converted into digital representations of the motif-specific label pattern. These data are then assembled de novo using data analysis software to re-create a whole-genome consensus map of the original genome. This genome map enables a variety of analyses, such as sequence finishing and SV detection.

Schema of principle of operation, relevant publications, collaborations, and operation protocols involving the BioNano Genomics system are listed on the website of the company (www.bionanogenomics.com). Opinions of users and the latest news can be found on the Genome website (www.genomeweb.com).

8.3 TECHNIQUES/SYSTEMS IN DEVELOPMENT

8.3.1 GnuBIO: Bio-Rad

The GnuBIO platform is a new desktop sequencer that incorporates all steps for DNA sequencing into a single system. Using emulsion microfluidics, GnuBIO has developed a scalable DNA sequencing reaction that encompasses all of the steps necessary for DNA sequencing inside picoliter-sized aqueous drops. Most other sequencing technologies require separate workflows for target selection, DNA amplification, DNA sequencing, and analysis; GnuBIO technology integrates complete workflow into a single instrument. The platform provides a single-step process that produces genomic results within hours. It is designed to meet the workflow of both research and clinical laboratories in applications from single genes to whole genomes. GnuBIO and its DNA sequencing platform was acquired by Bio-Rad in 2014 (for the schema of principle operations and more information, see www.gnubio.com).

8.3.2 Qiagen Platform

Several years ago, Qiagen announced a plan to develop and launch a DNA sequencing platform mainly for clinical applications, to add to its wide range of products for preparing and purifying nucleic acids. For this purpose they bought DNA sequencing technology from **SBH** (www.qiagen.com). The complete chain for DNA analysis, from consumables for DNA preparation to sequencing, and to data and functional analysis, would present a powerful combination. Some initial test results were conducted but the company has not yet launched the DNA sequencing platform, although all products for the other steps in the analysis chain, including the software for data and function analysis, have already been distributed by Qiagen, including its expertise in single cell analysis. The company's progress, applications, and projects can be followed on the website at www.qiagen.com.

8.3.3 Direct Genomics Single-Molecule Sequencing

Direct Genomics, a spinoff of the BGI institute in China, licensed its single-molecule sequencing technique from Caltech using a sequencing method first developed by Helicos Biosciences, Cambridge, Massachusetts. The **GenoCare** system will operate with chemistry and reagents from Helicos, but the technique and device were optimized and redesigned. The GenoCare is expected to have a throughput of around 10 Gb per run, with very high consensus accuracy (approaching 100% by 5× coverage), an initial read length of 35 bases, and 140 bases with four-color chemistry and improvements in hardware. The key improvement compared with the original Helicos technology is the use of total internal reflection fluorescence microscopy. The laser is reflected internally and illuminates only a thin region at the flow cell surface with captured DNA, resulting in a significantly improved signal-to-noise ratio. Another improvement is the use of the latest generation of charge coupled device cameras with higher sensitivity. The single-molecule technique does not require a

sample preparation module; DNA capture takes place directly on the flow cell surface. GenoCare has been designed for clinical applications in China, concentrating on sequencing of gene cancer panels and noninvasive prenatal screening and testing.

More information on the company's progress will be available on the BGI and Direct Genomics Web pages.

8.4 POTENTIAL FUTURE TECHNIQUES/SYSTEMS/ANALYSIS PLATFORMS

8.4.1 DNA Sequencing by Tunneling Currents

Besides the measurement of changes in ionic current as DNA passes through pores, there are efforts to measure electric signals perpendicular to the DNA molecule (in a technique called tunneling currents DNA sequencing) as it traverses the pore or channel. Expectations are that each base will show a sufficiently different tunneling current, and the base sequence will be deduced from the differences in the tunneling currents.

Depending on its electronic structure, each base may affect the tunneling current differently, allowing differentiation among bases during translocation through the pore or channel. The technique has the potential to sequence orders of magnitude faster than ionic current methods. Sequencing of several DNA oligomers and micro-RNA by the tunneling currents technique has been reported (Ohshiro et al., 2012; Di Ventra, 2013).

8.4.2 Mechanical Identification and Sequencing of Single DNA Molecule

The proof of concept was presented for a single-molecule platform that allows DNA identification and sequencing (Ding et al., 2012) based on the mechanical properties of DNA, as explained in the publication. An entirely novel interesting approach, the technique is not based on the detection of the fluorescent nucleotides but on DNA hairpin length. By mechanically pulling (in a magnetic tweezers system) on magnetic beads tethered by a DNA hairpin to the surface, the molecule can be unzipped and the two DNA strands mechanically separated. In this open state it can hybridize with complementary oligonucleotides, which transiently block the hairpin, rezipping when the pulling force is reduced. By measuring from the surface to the bead of a blocked hairpin, one can determine the position of the hybrid along the molecule with nearly single-base precision. The approach can be used to identify a DNA fragment of known sequence in a mixture of various fragments and to sequence an unknown DNA fragment by hybridization or ligation. The proof of concept of the technique has been demonstrated, but it may not be easy to develop it into a simple system for routine DNA sequencing applications, in competition with other techniques.

8.4.3 Sequencing With Mass Spectrometry

Matrix-assisted laser desorption ionization time-of-flight (TOF) mass spectrometry (MS) has been investigated specifically as an alternative method to gel electrophoresis for visualizing DNA fragments (Monforte and Becker, 1997; Hall et al., 2005). With this method, DNA fragments generated by chain-termination sequencing reactions are compared by mass rather than size and the DNA sequence is deduced. The mass of each nucleotide is different from the others, and this difference is detectable by MS. Single-nucleotide mutations in a fragment can be detected more easily with MS than by gel electrophoresis alone (Monforte and Becker, 1997; Edwards et al., 2005).

Early chain-termination and TOF MS methods demonstrated read lengths of up to 100 bp. To date, this appeared to be the limit for the measurable average read size. MS-based DNA sequencing may not be suitable for large de novo sequencing projects, just like chain-termination sequencing alone, but it has been used, e.g., in bacterial and forensic studies. A study used the short sequence reads and MS to compare single-nucleotide polymorphisms in pathogenic *Streptococcus* strains (Beres et al., 2010).

8.4.4 DNA Sequencing by Electron Microscopy

This approach directly visualizes the sequence of DNA molecules using electron microscopy. The detection is not straightforward because atoms composing the DNA are generally of low mass, and thus are not easily visualized by electron microscopic techniques. To be clearly visualized under an electron microscope (EM), DNA must be labeled with heavy atoms. In addition, specialized imaging techniques and aberration-corrected optics are beneficial for obtaining the resolution required to image the labeled DNA molecule. In theory, transmission electron microscopy DNA sequencing could provide extremely long read lengths, but the issue of electron beam damage to the DNA has not been entirely resolved, and the technology has not yet been developed commercially. Nevertheless the first identification of DNA bp within intact DNA molecules has been shown, which proved that modified bases containing atoms of increased atomic number were enzymatically incorporated. Direct visualization and identification of individually labeled bases within a synthetic 3272-bp DNA molecule and a 7249 bp viral genome were demonstrated (Bell et al., 2012; Mankos et al., 2012).

There are currently two companies working on the DNA sequencing technique using electron microscopy, employing different approaches. One company, **ZS Genetics**, published an initial proof-of-concept study in 2012 (Bell et al., 2012), and demonstrated that it was possible to label (with a single mercury atom) and identify one of four bases of DNA with an EM. The company claims that it is able to label and distinguish all four bases uniquely. The goal for DNA analysis with an EM is that it will enable long reads and high accuracy. Without the need for fragmentation, the DNA is first labeled; then the EM takes an image of it while analyzing and distinguishing among the different labels to read out the sequence. If confirmed in the future constructed device, such a strategy could enable reads of around 50,000 bp.

ZS Genetics uses transmission electron microscopy (TEM), in which electrons have about 1000 times more energy. In TEM, high-energy electrons are not scattered enough by the light elements that make up DNA, so the molecule cannot be visualized directly. The heavy metal labeling approach has several problems. One is that the labeling reaction might not go to completion, and thus some DNA bases may be missed and not detected. A further problem is that it is not yet possible to label only one type of base at a time because the labels are difficult to distinguish and they interfere with each other when they get too close. Thus the sequence is reconstructed from four separate images. In addition, high-energy electrons often damage the DNA, causing the labels to move around, so their position is no longer precisely located on the base.

Another start-up company, **Electron Optica**, and its collaborators have experimented and demonstrated in principal that sequencing of nonlabeled DNA by electron microscopy should be feasible. Technical approaches of the two companies differ in several aspects. Electron Optica will use low-energy electron microscopy (LEEM), in which electrons have low energies up to several hundred electron volts. Because of the lower beam energy in the LEEM device, damage to the DNA molecule will be smaller and it could reduce sequencing errors. Also, the low-energy electrons are slower and become scattered by the light elements of DNA (mostly carbon, oxygen, and nitrogen), enabling sufficient contrast to visualize DNA directly without the need to label the DNA with heavy metals, as was demonstrated in Mankos et al. (2012).

The company showed that the four base types provide sufficient contrast in LEEM for DNA sequencing. Electron Optica uses monochromatic aberration-corrected dual-beam LEEM, in which two beams illuminate the sample with low-energy electrons, and the reflected electrons compose an image. The LEEM technique has not been used widely in the life sciences because of the lower resolution compared with other types of EMs. Another reason for the less frequent use of the technique is that most biosamples are electrical insulators. As such, they charge up under the electron beam in the microscope, and the charge acquired by the sample affects the trajectory and lowers the resolution of low-energy electrons. To work around this problem, the company had to develop a charge-control technique.

Electron Optica is designing a new aberration corrector that can bring the resolution to about half a nanometer, which is needed to demonstrate that all four bases can be distinguished.

On four different kinds of substrates, and involving the use of advanced spectroscopy techniques such as X-ray photoelectron spectroscopy and Auger electron spectroscopy, the company achieved the contrast needed to distinguish among the individual bases. The difference in contrast among the bases is caused mainly by differences in the nitrogen content of the bases (Mankos et al., 2012).

The price of a finished sequencing microscope is estimated to be \$0.5−1 million. The platform would compete with other technologies on read length (the goal is to obtain base readings up to 50 Mb) as well as cost per base, accuracy (error rate of 10^{-6}), and throughput (one genome per day). Besides DNA sequencing, it should also be useful for RNA sequencing and detection of methylation on bases.

More details and information about progress of the techniques can be found on the websites of the companies (www.zsgenetics.com and www.electronoptica.com).

8.4.5 RNA Polymerase Sequencing Technique

This method (Pareek et al., 2011; Fujimori et al., 2012; Ansorge, 2009, 2010) is based on use of RNA polymerase (RNAP), which is attached to a polystyrene bead. One end of the DNA to be sequenced is attached to another bead, and both beads are placed in laser optical traps. RNAP motion during transcription brings the beads closer together, and as their relative distance changes, the change is recorded at single-nucleotide resolution. The sequence is deduced based on the four readouts with lowered concentrations of each of the four nucleotide types, similar to the Sanger method. Comparison is made between regions and sequence information is deduced by comparing known sequence regions to unknown sequence regions. Because of the special equipment needed, this technique may be useful in special applications and may not be easily developed into a wide technique competing with other available technologies.

8.4.6 Fluorescence Resonance Energy Transfer−Based and Raman Spectroscopy Approach for DNA Sequencing

A sequencing-by-synthesis approach that applies fluorescence resonance energy transfer (FRET) was proposed by

VisiGen Biotechnologies (www.visigenbio.com). In this technique a polymerase containing a donor fluorophore is used combined with four different acceptor fluorophores for the respective bases. Whenever the polymerase incorporates a nucleotide, an FRET signal is generated through the proximity of the donor and acceptor fluorophores. Sequence information is obtained consequently based on the specific labels for each base. Upon release of the pyrophosphate that contains the fluorophore, the FRET signal is quenched until the incorporation of the next base. The technique has been promising and was taken over by Life Technologies with no news concerning its eventual further development.

The use of surface-enhanced Raman scattering and tip-enhanced Raman scattering for DNA sequencing were described in a publication (Bailo and Deckert, 2008) as well as in a technical review (Treffer and Deckert, 2010). A test on a single RNA strand was reported as a promising step toward the development of a novel label-free single-molecule sequencing technique.

8.5 PERSPECTIVES FOR FUTURE APPLICATIONS AND DIAGNOSTICS TECHNIQUES

As the techniques and technologies are rapidly developing, DNA sequencing applications are spreading to all areas that are even remotely connected to DNA analysis. Examples are the control of vaccines and medicament purity in the pharma industry, where the detection of active viral components as well as inactive dormant viruses occurs. The control of claimed meat quality in the food processing industry, forensic analysis, fetal analysis from DNA circulating in the mother's blood, and the evolutionary biology of organisms are just a few other fields in which the technique has found application. Here, we focus on two emerging fields for which the latest DNA sequencing technology is of fundamental importance.

8.5.1 Precision Medicine: Personal Genomics

Precision medicine has gained increasing public attention and support (Collins and Varmus, 2015). As a result of technological advances that followed the Human Genome Project, the expectation emerged that the medical field would include the study of DNA, detecting genetically conditioned illnesses and predicting personalized responses to potential therapies. This new form of medical treatment would concentrate on a patient's specific needs for prevention, diagnosis, and treatment. Precision medicine has already revolutionized diagnosis and therapy for a number of cancers. Recognized as a 2013 Breakthrough of the Year by *Science Magazine*, a novel type of cancer immunotherapy has been developed. Genetically modified T-cells taken from blood and reengineered to detect a protein on cancer cells are used to fight specific cancer cells. Compared with chemotherapy, the molecular profile of individual patients will provide physicians with sufficiently specific details about the genetic condition to enable personalized treatment, minimizing adverse side effects and unnecessary exposure to less efficient treatments. Precision medicine will be help devise preventive approaches.

Currently, a person's genome can be sequenced at a cost of about $1000, and some companies already envision the price being lowered to around $100. This development indicates that DNA techniques providing individual molecular signatures may replace some traditional tools for diagnosis.

8.5.2 Single-Cell DNA Sequencing, Chips for Diagnostics

The isolation and analysis of single cells is a field of increasing interest. The goal of obtaining the sequence of a single DNA molecule from a single cell, ideally the entire length of a chromosome, without amplification and without destruction of the original DNA strand, as discussed in Ansorge (2009, 2010), has moved closer to reality. Rapid progress in single-cell techniques was described in detail in reviews (Wang and Navin, 2015; Macaulay and Voet, 2014). The technique enables the study of rare cells and will find applications in microbiology, neurobiology, development, immunology, and cancer research, as well as in the clinical environment.

Commercial devices for cell isolation are available (www.fluidigm.com), but most of the analysis still requires amplification steps. After single-cell isolation, advances in whole-genome and whole-transcriptome amplification have enabled the sequencing of low amounts of DNA and RNA present. The technique made it possible to study genomic and transcriptomic heterogeneity, because they develop in normal as well as diseased cells. It will also help to understand the evolution of genomic, epigenomic, and transcriptomic variations as they occur over a lifetime.

8.6 CONCLUSIONS

The diagnostics field has already profited from the advent of DNA sequencing technology, sample preparation, single-cell isolation techniques, DNA amplification schemes, and computer data analysis with improving software tools. Novel therapeutic approaches and public support for efforts in improvements in health care will no doubt give a steady push to develop and innovate the field further. Technology aims to lower the cost of the equipment and biochemicals involved, simultaneously increasing the

reproducibility, reliability, and simplicity of the techniques and protocols in operation.

The goal to be achieved, as formulated in Ansorge (2009, 2010), is still valid despite the great progress in DNA sequencing techniques since then. For genomic sequencing, and for the analysis of the ever more important structural genetic variations in genomes (e.g., CNVs for chromosomal translocations, inversions, large deletions, insertions, and duplications), if sequence read length on the original single DNA molecule could be increased to tens of thousands of bases and more in a relatively short time, it would be a great advantage. Ideally the goal would be the sequence determination of a whole chromosome from a single original DNA molecule, from a single cell.

So far, progress in this field has been encouraging, particularly in throughput and cost per base on Illumina systems, and single-molecule real-time approaches on Pacific Bioscience and Oxford Nanopore. Several new developments and improvements are still needed. The promise of benefits in health care and public interest in precision medicine will be the stimuli for continuous efforts in developments of technology, diagnostics, and therapies.

At present a human genome can be sequenced at a cost of about USD 1000. Some present developments indicate that the price could, in the foreseeable future be lowered down to USD 100. The DNA sequencing technique has the potential, with its precise determination of individual molecular signatures, to replace some traditional tools for diagnosis.

REFERENCES

Ansorge, W.J., 2009. Next generation DNA sequencing techniques and applications. New Biotechnol. 25 (4), 195–203.

Ansorge, W.J., 2010. In: Patrinos, G., Ansorge, W.J. (Eds.), Novel Next-Generation DNA Sequencing Techniques for Ultra High Throughput Applications in Bio-medicine, Molecular Diagnostics, second ed., p. 365.

Bailo, E., Deckert, V., 2008. Tip-enhanced Raman spectroscopy of single RNA strands: towards a novel direct-sequencing method. Angew. Chem. Int. Ed. 47, 1658–1661.

Bell, D.C., Thomas, W.K., Murtagh, K.M., Dionne, C.A., Graham, A.C., Anderson, J.E., Glover, W.R., October 2012. DNA base identification by electron microscopy. Microsc. Microanal. 18 (5), 1–5.

Beres, S.B., Carroll, R.K., Shea, P.R., Sitkiewicz, I., Martinez-Gutierrez, J.C., Low, D.E., McGeer, A., Willey, B.M., Green, K., Tyrrell, G.J., Goldman, T.D., Feldgarden, M., Birren, B.W., Fofanov, Y., Boos, J., Wheaton, W.D., Honisch, C., Musser, J.M., 2010. Molecular complexity of successive bacterial epidemics deconvoluted by comparative pathogenomics. Proc. Natl. Acad. Sci. 107 (9), 4371–4376.

Brock, A.P., et al., 2012. Accurate whole-genome sequencing and haplotyping from 10 to 20 human cells. Nature 487, 190–195.

Cherf, G.M., Lieberman, K.R., Rashid, H., Lam, C.E., Karplus, K., Akeson, M., 2012. Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. Nat. Biotechnol. 30, 344–348.

Collins, F.S., Varmus, H., 2015. A new initiative on precision medicine. N. Engl. J. Med. 372, 793–795.

Di Ventra, M., 2013. Fast DNA sequencing by electrical means inches closer. Nanotechnology 24 (34), 342501.

Ding, F., Manosas, M., Spiering, M.M., Benkovic, S.J., Bensimon, D., Allemand, J.F., Croquette, V., 2012. Single DNA molecule mechanical identification and sequencing. Nat. Methods 9, 367–372.

Drmanac, R., Sparks, A.B., Callow, M.J., Halpern, A.L., Burns, N.L., Kermani, B.G., Carnevali, P., Nazarenko, I., Nilsen, G.B., Yeung, G., Dahl, F., Fernandez, A., Staker, B., Pant, K.P., Baccash, J., Borcherding, A.P., Brownley, A., Cedeno, R., Chen, L., Chernikoff, D., Cheung, A., Chirita, R., Curson, B., Ebert, J.C., Hacker, C.R., Hartlage, R., Hauser, B., Huang, S., Jiang, Y., et al., 2010. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science 327 (5961), 78–81.

Drndic, M., 2014. Sequencing with graphene pores. Nat. Nanotechnol. 9, 743.

Edwards, J.R., Ruparel, H., Ju, J., 2005. Mass-spectrometry DNA sequencing. Mutat. Res. 573 (1–2), 3–12.

Fujimori, S., Hirai, N., Ohashi, H., Masuoka, K., Nishikimi, A., Fukui, Y., Washio, T., Oshikubo, T., Yamashita, T., Miyamoto-Sato, E., 2012. Next-generation sequencing coupled with a cell-free display technology for high-throughput production of reliable interactome data. Sci. Rep. 2, 691.

Garaj, S., et al., 2010. Nature 467, 190–193.

Hall, T.A., Budowle, B., Jiang, Y., Blyn, L., Eshoo, M., Sannes-Lowery, K.A., Sampath, R., Drader, J.J., Hannis, J.C., Harrell, P., Samant, V., White, N., Ecker, D.J., Hofstadler, S.A., 2005. Base composition analysis of human mitochondrial DNA using electrospray ionization mass spectrometry: a novel tool for the identification and differentiation of humans. Anal. Biochem. 344 (1), 53–69.

Kumar, T., Chien, H., Balijepalli, R., Li, R., Reiner, Kasianowicz, J., 2012. PEG-labeled nucleotides and nanopore detection for single molecule DNA sequencing by synthesis. Sci. Rep. 2, 2012 (Nature Publications group), Article number 684.

Laszlo, A.H., Derrington, I.M., Ross, B.C., Brinkerhoff, H., Adey, A., Nova, I.C., Craig, J.M., Langford, K.W., Samson, J.M., Daza, R., Doering, K., Shendure, J., Gundlach, J.H., 2014. Decoding long nanopore sequencing reads of natural DNA. Nat. Biotechnol. 32, 829–833.

Liu, K., Feng, J., Kis, A., Radenovic, A., 2014. ACS Nano 8, 2504–2511.

Macaulay, I.C., Voet, T., 2014. Single cell genomics: advances and future perspectives. PLoS Genet. 10 (1), e1004126. http://dx.doi.org/10.1371/journal.pgen.1004126.

Mankos, M., Shadman, K., N'diaye, A.T., Schmid, A.K., Persson, H.H., Davis, R.W., November 2012. Progress toward an aberration-corrected low energy electron microscope for DNA sequencing and surface analysis. J. Vac. Sci. Technol. B Nanotechnol. Microelectron. 30 (6), 6F402. Epub 2012 October 26.

Manrao, E.A., Derrington, I.M., Laszlo, A.H., Langford, K.W., Hopper, M.K., Gillgren, N., Pavlenok, M., Niederweis, M., Gundlach, J.H., 2012. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. Nat. Biotechnol. 30, 349–353.

Merchant, C.A., et al., 2010. Nano Lett. 10, 2915–2921.
Monforte, J.A., Becker, C.H., 1997. High-throughput DNA analysis by time-of-flight mass spectrometry. Nat. Med. 3 (3), 360–362.
Ohshiro, T., Matsubara, K., Tsutsui, M., Furuhashi, M., Taniguchi, M., Kawai, T., 2012. Single-molecule electrical random resequencing of DNA and RNA. Sci. Rep. 2, 501. http://dx.doi.org/10.1038/srep00501.
Pareek, C.S., Smoczynski, R., Tretyn, A., 2011. Sequencing technologies and genome sequencing. J. Appl. Genet. 52 (4), 413–435.
Schneider, G.F., et al., 2010. Nano Lett. 10, 3163–3167.
Treffer, R., Deckert, V., 2010. Recent advances in single-molecule sequencing. Curr. Opin. Biotechnol. 21, 4–11.
Wang, Y., Navin, N.E., 2015. Advances and applications of single cell sequencing technologies. Mol. Cell 58 (4), 598–609.

Chapter 9

Advanced Personal Genome Sequencing as the Ultimate Diagnostic Test

B.A. Peters[1,2], S. Drmanac[1,2], J.S. Liu[1], X. Xun[1,2] and R. Drmanac[1,2]
[1]Complete Genomics, Inc., Mountain View, CA, United States; [2]BGI-Shenzhen, Shenzhen, China

9.1 INTRODUCTION

Diagnostic tests are a critical component of an advanced health care system and are given to patients at practically every stage of disease management. Most modern diagnostic tests examine one or a small number of specific markers that have been studied and shown to correlate with disease states (see also Chapter 2). In the field of DNA diagnostics this paradigm has persisted. However, our knowledge is rapidly expanding and the model of developing a test for and examining one marker at a time is quickly becoming outdated. In this chapter we will explore why we need to change the current strategy and shift to whole-genome sequencing (WGS), which we consider to be the ultimate genetic test.

The development of the polymerase chain reaction (PCR) (Saiki et al., 1985) and the Sanger sequencing based semiautomated DNA sequencer (Smith et al., 1986) in the mid 1980s fundamentally revolutionized the field of molecular diagnostics. Before this, only a few diagnostic tests existed to identify specific variants causing diseases such as thalassemia and cystic fibrosis. These new tools allowed for the simple development of gene tests that could examine multiple variants at one time and created a platform for the discovery of additional disease-associated variants. By 1995 it was clear that structure and guidance were needed for this emerging field, thus heralding the formation of the Association of Molecular Pathology. The Human Genome Project, which culminated in the completion of the reference sequence of the human genome in 2003, allowed for the identification and mapping of almost every gene in the human genome and set the stage for an explosion in the number of disorders with a specific gene test available (Fig. 9.1). As of 2016, almost 5000 were available (www.genetests.org). At the same time, improvements in DNA sequencing technology, especially in the massively parallel "next-gen" sequencing, resulted in a progression in comprehensiveness from one to a panel of mutations examined and from whole-gene sequencing (e.g., *BRCA1*) to whole-exome sequencing. Now, with the most recent advances in WGS, there are almost no limits to our ability to understand and diagnose the genetic basis of human disease.

9.1.1 Why Individual Whole-Genome Sequencing Is Important

Within each of us lies a program for the development and adaptive functioning of all of our tissues. This genome program is coded by the 3 billion DNA letters transferred to us from each parent. It is the ultimate genetic code; there is no "deeper" level to be discovered. As such, it is sufficiently predictive for our future body and mind functioning and enables an early (e.g., prenatal) presymptomatic baseline test, although predictions such as heart failure or cancer predisposition are frequently conditional owing to environmental and stochastic factors such as epigenetic states that become frozen within a population of cells creating an "epigenetic lock-in." It can also predict which drugs are potentially useless for each of us or, even more important, those that could be deadly (Mizzi et al., 2014), such as in the case of a 2-year-old boy who died from potential complications of a codeine overdose and was later discovered to have a cytochrome P-450 2D6 fast-metabolizer phenotype that would have precluded him from treatment with that drug (Ciszkowski et al., 2009). Importantly, after WGS analysis and proper interpretation, an individual genome program is potentially "modifiable"

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00009-2
Copyright © 2017 Elsevier Ltd. All rights reserved.

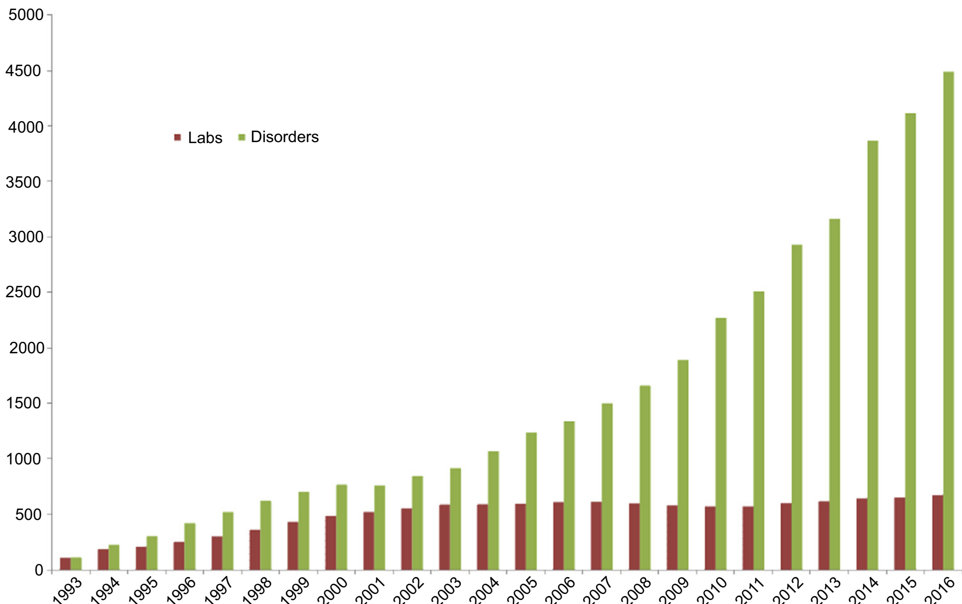


FIGURE 9.1 Increasing number of disorders with gene tests available. Since 1993 the number of gene tests for specific disorders has increased from less than 500 to close to 5000. *This figure was reproduced with permission from www.GeneTests.com from data on their website. www.GeneTests.com aggregates worldwide testing information from over 678 laboratories and 1067 clinics.*

through personalized interventions such as body and mind exercises, diet, supplements, and medications. In the near future we may be able to edit the genome of our stem cells to create healthier tissues. In essence, personal WGS will help to inform us of all aspects of our life and the infrastructure to enable population-scale WGS may become as important as that which enables the Internet. Many industries related to human health and well-being will be supported by this new infrastructure. Sequencing technologies may change but the need to read, store, and interpret millions if not billions of whole human genome sequences per year will exist in perpetuity.

9.2 ADVANCED WHOLE-GENOME SEQUENCING

In 1965, Gordon E. Moore, cofounder of Intel, made the observation that the number of transistors in an integrated circuit doubles approximately every year (later revised to every 2 years in 1975), resulting in the cost of a chip with a fixed number of transistors dropping approximately half over that period (Moore, 1965). This rapid pace of development resulted in the remarkable reduction in the cost and size of integrated circuits such that a modern pocket-sized smartphone (iPhone 5, about $350) has four orders of magnitude more computing power and lower cost than the supercomputers that filled an entire floor used by NASA during the Apollo program (IBM System/360 Model 75, $3.5 million). Known as “Moore’s law,” this pattern has become the benchmark by which all technologies are compared. Genome sequencing is no exception; since the conclusion of the $3 billion, 13-year human genome project in 2003, the price of sequencing all 3 billion base pairs of the human genome has rapidly dropped to the current cost of about $1000, considerably outpacing Moore’s law (Fig. 9.2). Concurrent with the huge drop in cost has been the massive reduction in time, from 13 years to sequence the reference genome to 26 h today for a personal genome (Miller et al., 2015).

Similar to the computer microchip, increasing the number of features that can be analyzed in parallel has driven much of this rapid reduction in cost and time. In DNA analysis, this so-called massively parallel sequencing, pioneered a quarter century ago (Drmanac and Crkvenjakov, 1990), involves breaking the genome into many fragments, distributing them across an array, and analyzing them simultaneously. The latest generation of high-throughput sequencers now contains chips with billions of these features and can analyze multiple whole human genomes in a day. For this process to work, it requires each location on the array to be a fragment from only one location in the genome. An obvious solution is to have a single DNA molecule per location. However, this creates the problem of creating a sufficient signal from each fragment to allow for accurate base sequencing in a cost-effective manner. The solution has been clonal amplification of each fragment, typically by PCR on micron-sized beads in an emulsion (Drmanac and Crkvenjakov, 1990) or directly on the array support by in situ PCR (Bentley et al., 2008). An alternative (Drmanac et al., 2007, 2010b)

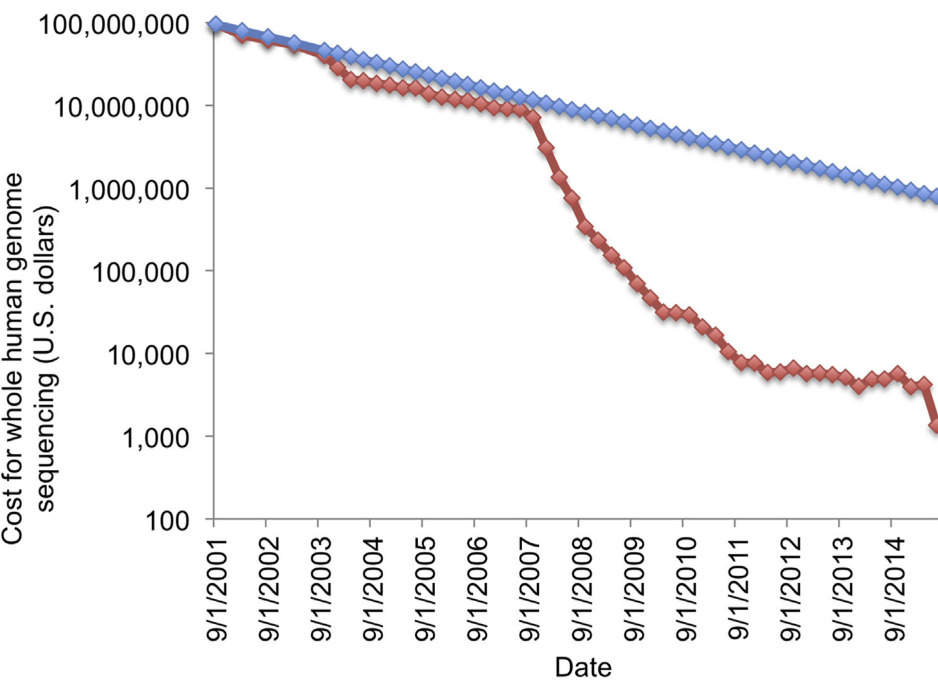


FIGURE 9.2 Moore's law versus DNA sequencing. Named after the cofounder of Intel, Moore's law is the observation that the number of transistors in an integrated circuit doubles approximately every 2 years (Moore, 1965). This results in the cost of electronics dropping by approximately half in that period. Plotting the cost of sequencing (*red* Dark Gray print in versions) against the rate of cost reduction based on Moore's law (*blue* Light Gray print in versions) demonstrates how rapidly sequencing technology advanced from September 2001 to February 2015. *Reproduced from Wetterstrand, K.A. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Available at: www.genome.gov/sequencingcostsdata.*

developed by Complete Genomics (Mountain View, CA) uses a rolling circle replication process to generate hundreds of clonal copies of each DNA fragment as a concatamer that fold into DNA nanoballs (DNBs) about 200 nm in diameter (Fig. 9.3). This process can be done in a single homogeneous reaction resulting in billions of these DNBs. The inherent negative charge of these DNA particles allows them to adhere to a high-density nanoarray with a regular pattern of positively charged 200-nm spots spaced about 500−1000 nm center to center, which makes this a highly effective, low-cost method for creating a nanoarray with more than a billion unique clonal spots. Because of their size and related physical and chemical properties only one nanoball can fit per "sticky" spot, resulting in over 95% of spots occupied with a single DNB. Furthermore, controlled multiple displacement amplification can be performed in situ on the arrayed DNBs to generate over 1000 template copies per 200-nm spot, providing an order of magnitude higher signal density than in PCR clusters (Drmanac et al., 2007). Such dense nanoarrays consume less sequencing reagents and allow highly efficient use of digital imaging computer chips to the theoretical maximum of imaging a DNA spot with a single camera pixel (Peters et al., 2012).

Although these achievements have arrived at a mind-boggling rate, much remains to be improved (Table 9.1). Humans are diploid; that is, we inherit a set of homologous chromosomes from each parent, but information about parental variants is completely lost during the fragmenting of the genome required by most highly parallel sequencing systems. This kind of information can be critical in situations where an individual has two different detrimental variations in a single gene. If the two variants are on different homologous chromosomes, the potential impact of this could be a loss of function of that gene product. Furthermore, although error rates are low with greater than 50× read coverage and proper variant calling (one error in approximately 1 million bases), they still cause several thousand false "variants" in each genome that can cloud genome interpretation. In addition, current DNA sequencing misses many regions of the genome, and because short sequences must be mapped back to the reference genome, important individual variation within each personal genome is lost. Furthermore, assigning variants to specific members of gene families including pseudogenes and recent segmental duplications is often impossible. These short comings have been understood for many years and despite efforts in developing advanced mate-pair libraries (Bentley et al., 2008; Drmanac et al., 2007, 2010a; Kim et al., 2009; Ahn et al., 2009; Ley et al., 2008; McKernan et al., 2009; Pushkarev et al., 2009; Wang et al., 2008; Wheeler et al., 2008) and algorithms for assembling such short reads using local de novo assembly and variant calling based on Bayesian probabilities

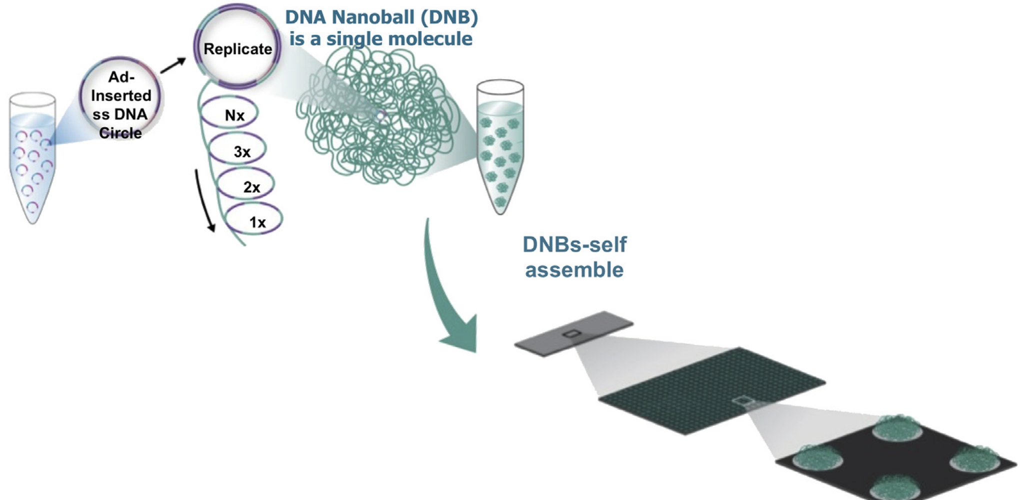


FIGURE 9.3 Massively parallel analysis: DNB arrays. Complete Genomics' proprietary DNB arrays are a powerful example of how massively parallel analysis is performed. Creation of an array starts with rolling circle replication of single-stranded circles. This can be performed on billions of circles in a single small volume (a few hundred microliter reaction) resulting in billions of DNBs that consist of single-stranded tandem copies of fragments of genomic DNA interrupted by known adapter sequences. These long single-stranded molecules form negatively charged nanoballs (DNBs). This negative charge helps keep the DNBs from aggregating and also provides a simple method for arraying the DNBs. The DNBs are spread across a silicon chip with a patterned array of positively charged spots spaced approximately 500−1000 nm from center to center, resulting in close to 100% occupancy of all spots on the array with a single DNB.

(Carnevali et al., 2012), these problems remain. Finally, most current technologies require thousands to millions of cells' worth of DNA, eliminating many sample types from analysis (Table 9.1).

But it is not hopeless; as in all things in technology there are solutions. Many of the data quality problems described in Table 9.1 are the result of breaking the genome into small fragments (approximately 0.5−5 kb). One approach to ameliorate this problem of massively parallel sequencing is the concept of co-barcoded reads, first described in 2006 by Drmanac and since used by multiple groups (Peters et al., 2012, 2015, 2014; Amini et al., 2014; Kuleshov et al., 2014; Kaper et al., 2013; Zheng et al., 2016) to improve genome sequencing quality. The idea is simple: Add the same unique barcode to each short fragment derived from a single long genomic DNA molecule. After DNA and barcode sequencing, the barcode allows one to determine which short (e.g., 100 bases) sequence reads came from a particular long genomic DNA molecule or small pools of such DNA molecules. In the most simple design this can be performed using 96−384 barcodes (Fig. 9.4) on 96−384 pools of a few thousand long genomic DNA molecules (>50 kb) in a process called long fragment read (LFR) technology (Peters et al., 2012). Using this information allows for determination of the order of variants from each parental chromosome (the phase or haplotypes) as well as significant error correcting down to one error in 1 billion bases (Peters et al., 2012, 2015). However, 384 barcodes is not sufficient to fix all of the current problems. Scaling the number of barcodes to over 10,000, possibly to a million or more, has the potential to generate near-perfect genome assembly (Peters et al., 2014). This large number of barcodes can be implemented in oil nano-drops proposed several years ago (Drmanac et al., 2010a) and implemented (Zheng et al., 2016), or potentially in a single tube in which long DNA fragments interact with millions of microbeads containing clonal barcodes 10−15 bases in length (Drmanac et al.,

TABLE 9.1 Limitations of Current Whole-Genome Sequencing
The order of variants on parental chromosomes is not retained (i.e., limited haplotype information).
There is difficulty mapping repetitive sequences.
There are high error rates.
Thousands to millions of cells are required excluding many sample types from analysis (e.g., in vitro fertilized embryo biopsies, blood circulating tumor and fetal cells, microbiopsies, etc.).

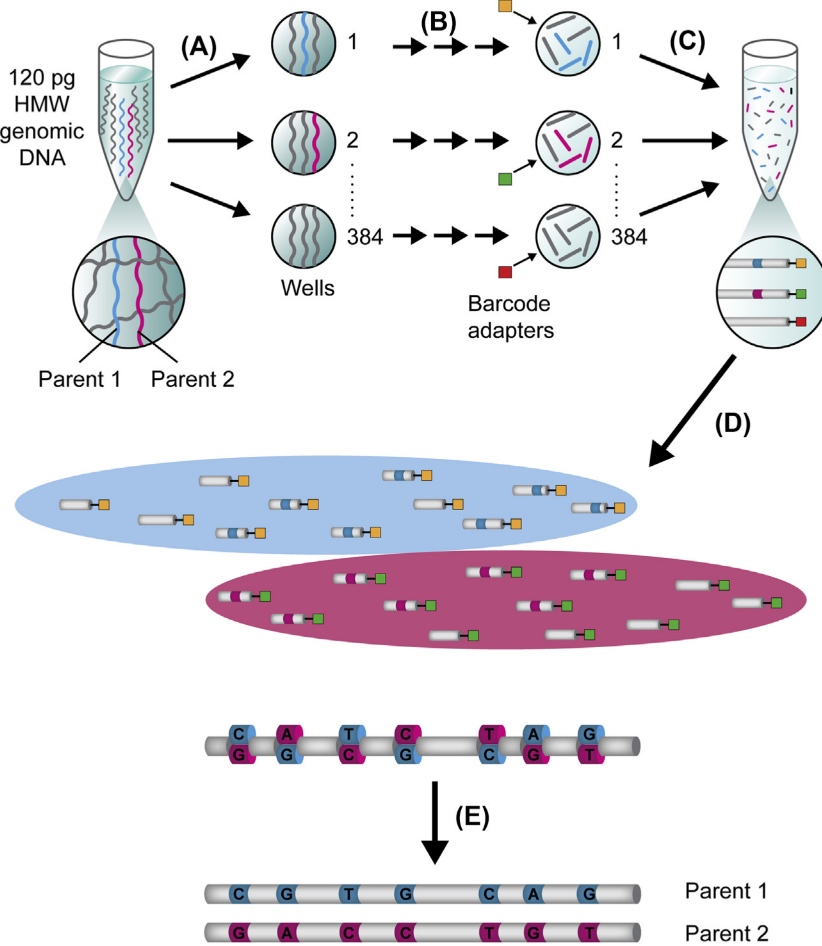


FIGURE 9.4 Long fragment read (LFR) technology. The LFR process starts with about 120 pg of high−molecular weight (HMW) genomic DNA or directly from as few as five cells. Long genomic DNA molecules are denatured to single-strand DNA and dispersed across a 384-well plate. The result is that much less than the amount of DNA contained within a single genome is found in each well. Consequentially the likelihood of overlapping fragments of DNA from homologous chromosomes of each parent occurring in a single well becomes low. Through amplification, fragmenting, and ligation of barcodes, these individual long fragments that are derived from a single parent can be reassembled in silico. This allows for the determination of which variants were inherited as long blocks of DNA from each parent.

2013). Coupled with advanced computational algorithms, this type of data would allow phased and reference-unbiased de novo whole-genome assembly (Peters et al., 2014). This would resolve segmental duplications, including proper assembly of active and pseudogenes, and could possibly measure telomere lengths for each of the 92 chromosomal ends of the human genome. To resolve any remaining difficult regions in the genomic sequence, this advanced co-barcoding could be combined with megabase single-molecule mapping (Mak et al., 2016) and continuous 10-kb or longer reads (Levene et al., 2003).

9.2.1 Advanced Whole-Genome Sequencing as the Ultimate First-Line Diagnostic Test of Sick Children

In our current medical system, when a child is born with a serious unknown disease the process of trying to determine what is causing the observed symptoms begins. According to one study of leukodystrophies, 20 tests were performed on average before a diagnosis was made (Richards et al., 2015). This convoluted path of various specific clinical tests has led to the use of the term "diagnostic odyssey" as a description. Often this accumulation of tests exceeds $15,000 and can take years with no satisfactory answer. As mentioned, our genome represents the deepest level of instruction for how to build each of us. As such, complete analysis of that code can allow for the identification of places where the instructions have a mistake, ultimately leading to a disease. Studies have shown that thousands of genes (O'Roak et al., 2012; Sanders et al., 2012; Michaelson et al., 2012; De Ligt et al., 2012; Veltman and Brunner, 2012; Epi et al., 2013; Yang et al., 2013; Iossifov et al., 2014; Al Turki et al., 2014; De Rubeis et al., 2014; Gilissen et al., 2014) may result in disease when a mistake, typically a

random de novo mutation in the sperm or egg, is made in the coding or regulatory sequence of one of them. In addition, as mentioned in the Introduction, there are currently over 5000 genes purported to be associated with a genetic disease for which single-gene testing is available. Beyond that, there are likely to be hundreds to thousands of genes that when carrying specific variations or de novo mutations can modify the phenotypes of diseases caused by mutations in other genes (Hamilton and Yu, 2012). In light of these realizations, it is increasingly clear that looking at a few spots is not only unlikely to find answers but is also a waste of time and money. The future approach will be to look at everything at once through WGS (or at least to look at all coding through whole-exome sequencing) as the first diagnostic test. As the cost of sequencing has become more affordable it is starting to be used in some of these cases and the results have been extremely positive (Vissers et al., 2016). In a high-quality (80× coverage) WGS study of children with intellectual disability and developmental delay, 62% of patients had a conclusive diagnosis from whole-genome data compared with only 12% using more conventional microarray technology (Gilissen et al., 2014). Additional studies demonstrated similar stories of success (Table 9.2). In most cases, these studies detected random de novo genetic alterations of single-base substitutions to short deletions or insertions.

In the case of a child with disease, the whole genomes or exomes of both parents are typically sequenced as well. This is to help reduce errors (high-quality base calls are required to be made in all three subjects at each called position) and identify de novo mutations that are often found to be the causative variant in the children of parents without disease or a history of disease within their families. It is also helpful in cases where there are two different recessive variants in a gene, and it is then critical to know whether each copy of the gene harbors a variant (compound heterozygous inactivation) or whether both variants are found in one allele. However, as our knowledge grows about what variants cause disease, it may become unnecessary to know whether something is de novo; the nature of the variation and the gene in which it occurs will be sufficient to indicate whether it causes disease. A study by Samocha et al. (2014) demonstrated this concept by describing a set of genes in which variation is restricted, and with such coding de novo mutations are more likely to be disease causing. Nonetheless, knowing the phase of variants and mutations will still be critical to identify disease-causing compound heterozygous events. As mentioned earlier, haplotyping a child's genome with almost error-free mutation calls can be determined directly through methods such as LFR. Improved knowledge of variant impact and employing LFR will ultimately obviate the need for parental sequencing in these cases and result in additional cost savings.

TABLE 9.2 Diagnostic Yield of WGS/WES in Published Studies

Year	Study	Disease	Methods	Diagnostic Yield (%)	Subjects, n
2011	O'Roak et al.	ASD	WES	25	20
2012	O'Roak et al.	ASD	44-gene panel	1	2446
2012	De Ligt et al.	DD/ID	WES	16	100
2013	Jiang et al.	ASD	WGS	50	32
2013	Yang et al.	Multiple disorders	WES	25	250
2014	De Rubeis et al.	ASD	WES	6.7	3871
2014	Iossifov et al.	ASD	WES	30	2517
2014	Gilissen et al.	DD/ID	WGS	42	50
2015	Deciphering Developmental Disorders study	DD/ID	WES	28	1133
2015	Yuen et al.	ASD	WGS	42	170
2015	Tammimies et al.	ASD	WES	8.4	95
2015	Shaheen et al.	DS	WES	90	31
2016	Monroe et al.	DD/ID	WES	29.4	17

ASD, autism spectrum disorder; *DD/ID*, developmental delay/intellectual disability; *DS*, dysmorphology syndromes; *WES*, whole-exome sequencing; *WGS*, whole-genome sequencing.

9.2.2 Advanced Whole-Genome Sequencing as Ultimate Prenatal/Preimplantation Test to Reduce Prevalence of Genetic Diseases

Worldwide, approximately 134 million babies are born each year, 7.9 million (about 6%) of whom are born with a serious birth defect or early-onset disease (Christianson et al., 2006). Of those children, 3.3 million die within the first 5 years of life and another 3.2 million may survive but be severely disabled for the remainder of their lives. Even in developed countries the percentage of children born with birth defects is high (approximately 3% in the United States) and 20% of those children die within the first year of life (Matthews et al., 2015). Many of these birth defects are the result of inherited variants, large-scale chromosomal gains or losses, or de novo mutations, all of which can be detected with WGS. In addition to birth defects, severe genetic diseases such as autism and other rare diseases affect up to 10% of the population (www.nlm.nih.gov/medlineplus/rarediseases.html).

In developed countries the infrastructure is available to screen prospective parents for a handful of recessively inherited disorders and probably contributes to some of the reduction in the number of birth defects compared with developing nations. However, whereas most birth defects and rare diseases are genetically based, many are not inherited but instead are the result of random events that occur in the germline of the parents or possibly early in the development of the embryo. Furthermore, these random defects have been shown to increase as the age of the parents increases (Figs. 9.5 and 9.6) (Munne et al., 1995; Crow, 2000; Hassold and Hunt, 2009; Kong et al., 2012; Jiang et al., 2013; Gratten et al., 2014; Francioli et al., 2015; Steiner et al., 2015; Wong et al., 2016). In these cases prenatal or preimplantation, if in vitro fertilization (IVF) is performed, screenings are the only method to detect these defects. In both cases there are tests that can detect chromosomal copy number changes.

An example of one of these noninvasive prenatal tests (NIPTs) for prenatal screening, called NIFTY, was developed by BGI in China and is based on collecting cell-free DNA (cfDNA) from a small sample of maternal blood (10 mL). During pregnancy, cfDNA fragments originating from both the mother and the fetus are present in the maternal blood circulation. By sequencing about 10 million fragments (base coverage representing about 10% of a single whole genome) of this cfDNA from the maternal blood, chromosomal abnormalities in the fetus can be detected. If aneuploidy is present, small increases or decreases in the counts of the affected fetal chromosome will be detected. Using WGS technology and four different proprietary bioinformatics analysis pipelines, the NIFTY test is able to analyze data across the entire genome and compare chromosomes in the tested sample against optimal reference chromosomes to determine the presence of genetic abnormalities accurately. Over 500,000 cost-effective NIFTY tests were performed by BGI in 2015 on the first China Food and Drug Administration–approved sequencing test based on Complete Genomics' DNB nanoarray technology. NIFTY was found by an independent study to have the highest NIPT sensitivity and specificity rates (Cheung et al., 2015).

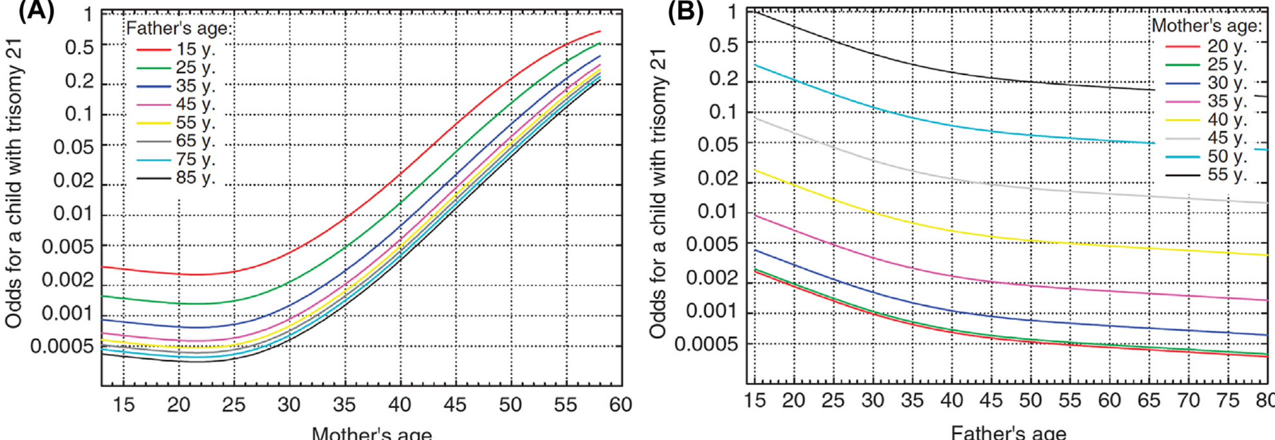


FIGURE 9.5 Aneuploidy is more prevalent in older mothers. (A) It has long been known that chromosomal abnormalities are more common in the embryos and newborns of older mothers (Munne et al., 1995; Crow, 2000; Hassold and Hunt, 2009), as shown by an increasing risk of giving birth to a child with trisomy 21 with increasing maternal age. (B) However, this association is not seen in older fathers, and indeed studies suggest that younger fathers may have a slightly higher risk of conceiving a child with trisomy 21 (Steiner et al., 2015). *These graphs were reproduced from Steiner, B., Masood, R., Rufibach, K., Niedrist, D., Kundert, O., Riegel, M., Schinzel, A., 2015. An unexpected finding: younger fathers have a higher risk for offspring with chromosomal aneuploidies. Eur. J. Hum. Genet. 23, 466–472, with permission from the Nature Publishing Group (License No. 3,879,540,189,717).*

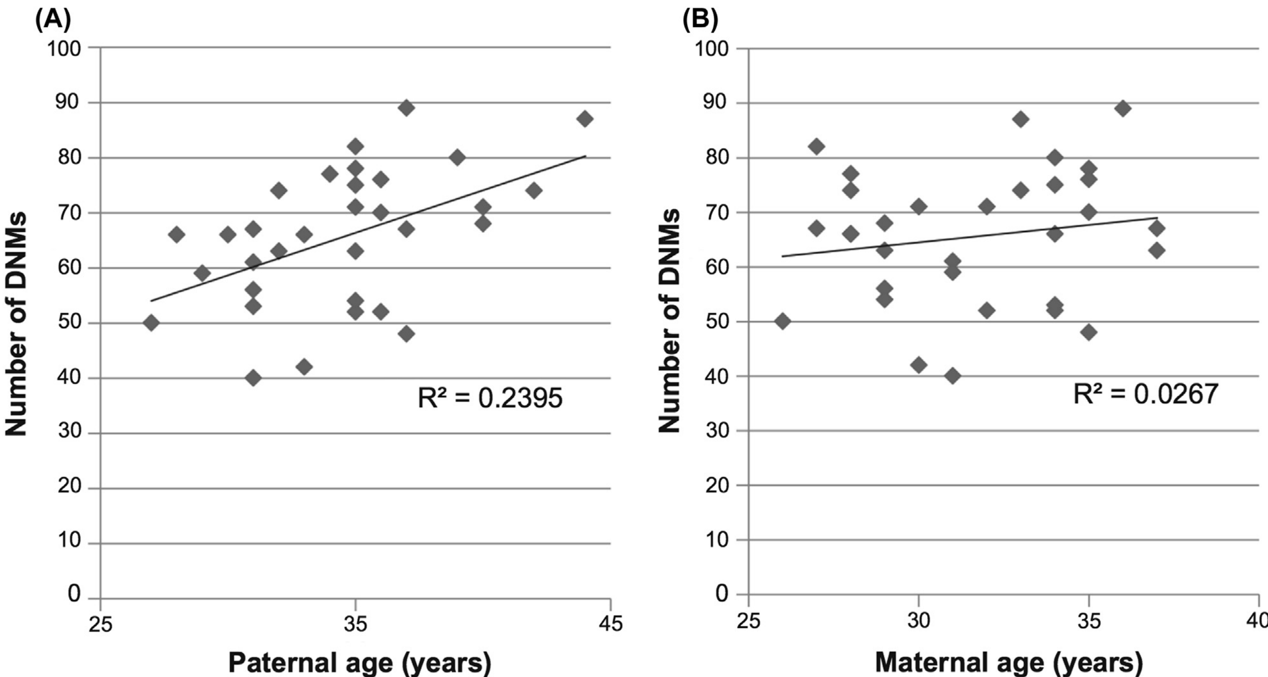


FIGURE 9.6 De novo mutations (DNMs) are more prevalent in older fathers. (A) A number of studies (Kong et al., 2012; Jiang et al., 2013; Gratten et al., 2014; Francioli et al., 2015; Wong et al., 2016) demonstrated that older fathers contribute a larger number of de novo mutations to their offspring. (B) But older mothers do not. *These graphs were reproduced from Jiang, Y.H., Yuen, R.K., Jin, X., Wang, M., Chen, N., Wu, X., Ju, J., Mei, J., Shi, Y., He, M., Wang, G., Liang, J., Wang, Z., Cao, D., Carter, M.T., Chrysler, C., Drmic, I.E., Howe, J.L., Lau, L., Marshall, C.R., Merico, D., Nalpathamkalam, T., Thiruvahindrapuram, B., Thompson, A., Uddin, M., Walker, S., Luo, J., Anagnostou, E., Zwaigenbaum, L., Ring, R.H., Wang, J., Lajonchere, C., Wang, J., Shih, A., Szatmari, P., Yang, H., Dawson, G., Li, Y., Scherer, S.W., 2013. Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. Am. J. Hum. Genet, 93, 249–263, with permission from Elsevier (License No. 3,879,491,391,474).*

Although tests such as NIFTY are a big step forward in detecting large-scale defects in the genome, they do not have the resolution to detect potentially disease-causing variants smaller than several megabases. In these cases new technologies are needed and, most likely owing to the small amount of fetal DNA found in the mother's blood (<10% of the total DNA), new sources of genetic material from the fetus will need to be used. An intriguing possibility is to use maternal haplotypes (e.g., determined by LFR-based WGS) and deep molecular-barcode–enabled targeted sequencing of cfDNA to detect both maternal and paternal inherited as well as de novo mutations in approximately 2000 genes causing the most serious diseases.

One active area of research is collecting fetal cells that circulate in the maternal bloodstream. Currently there are no technologies that can collect and enrich these cells to a consistent level in which more than 10% of cells are from the fetus. Furthermore, because these cells are rare, it is unrealistic to expect to collect much more than 10 cells per mother. In this case, sequencing such a small number of cells will require a large amount of amplification to generate enough material for most genome sequencing platforms. This large amount of amplification will result in thousands of false-positive polymerase errors, which appear as de novo mutations during the data analysis stage, masking the fewer than 100 true de novo mutations (Peters et al., 2015). However, methods are currently available, such as LFR, that allow for the analysis of 10 cells and removal of the vast majority of these misleading false-positive errors (Peters et al., 2012, 2015).

Another possible source of material is from the cells collected during amniocentesis. Almost all of these cells are from the fetus and millions can be collected, making enrichment a nonissue. However, amniocentesis is an invasive procedure and is less suitable for use in all pregnancies. That said, often older mothers (aged over 35 years) are referred for this procedure because of the increased rate of chromosomal abnormalities and birth defects such as Down syndrome in this age group. In these cases, WGS could provide chromosomal copy number data as well as additional small variant information, ultimately improving detection of all birth defects.

IVF is another area for which advanced comprehensive genetic testing could have a large impact. It has been estimated that currently 350,000 babies are born yearly through IVF (De Mouzon et al., 2009, 2012; Centers for Disease Control and Prevention, 2011; ESHRE, 2012). That number is expected to rise, because advanced maternal age is associated with decreased fertility rates and women

in developed countries continue to delay childbirth to later ages. Although that is much less than 1% of all births worldwide, it still represents many babies each year and the rate of birth defects in this group is approximately the same as in the general population. In addition, a large subpopulation of parents aged over 35 years who, owing to difficulty in conceiving naturally, choose to conceive through IVF are expected to have a higher likelihood of giving birth to children with birth defects. As with prenatal screening, the current methods for detecting genetic defects in the embryo are limited and miss many of the de novo events that cause birth defects. In the case of preimplantation testing for IVF, cell purity is not an issue, but the number of cells is. Testing typically involves the biopsy of one cell from a 3-day embryo or the more preferred method, owing to improved implantation success rates (Scott et al., 2013), of up to 10 cells from a 5- to 6-day blastocyst-stage embryo. As with fetal cells isolated from the blood, polymerase errors incorporated during amplification are a potential issue and strategies such as LFR can be used to avoid these false-positive de novo—like errors. In fact, this method has been demonstrated to work effectively on IVF biopsies (Peters et al., 2015). Reading almost all of the approximately 6 billion bases in the embryonic genome, LFR was able to detect over 80% of the de novo mutations and call fewer than six errors. Furthermore, most of the detected de novo mutations were phased, allowing for the proper diagnosis in imprinted genes (Schaaf et al., 2013).

9.2.3 Advanced Whole-Genome Sequencing as Ultimate Cancer Profiling for Selecting Optimal Cancer Therapy

Many cancer therapeutics have been developed toward specific somatic alterations found in cancer, which makes it critical to understand what mutations are present in each tumor before starting therapies. This has led to the relatively rapid implementation of genetic analysis methods for cancer patients (e.g., Foundation One, PGDx, and BGI's TumorCare). Most of these focus on a select group of genes commonly found to harbor cancer mutations. Although these methods have helped cancer patients receive the correct drugs, they are blind to many large chromosomal changes as well as many other somatic mutations that could be targeted to improve outcomes. Like all of the other examples we have talked about in this chapter, an unbiased whole-genome analysis of every patient's tumor would allow for the detection of almost all important and targetable variants and allow for an optimal personalized therapy.

Cancer presents multiple problems for whole-genome sequence analysis. First, acquisition of genetic material for sequencing can be challenging. In cases where the tumor is accessible for biopsy or is surgically removed, obtaining a sufficient amount of DNA is typically not a problem. That said, many tumors are heavily infiltrated with normal cells and can result in a low percentage of tumor cells in the material to be analyzed. In addition, most surgically removed tumors or biopsies are often fixed in formalin and embedded in paraffin. This dramatically reduces the quality of the DNA isolated from this material and can increase the number of false-positive variants in the ensuing genome analysis. In the case where there is no material directly from the tumor, genetic material of the cancer can be isolated from the bloodstream in a manner analogous to collecting fetal material from the mother's bloodstream. Both tumor cfDNA (ctDNA) and free circulating cancer cells (CTCs) can be collected from the blood. In the case of ctDNA the percentage of material that is derived from the tumor varies widely and is frequently low (less than 1%) (Sausen et al., 2014). Without enrichment this would require about 100× more sequence reads to generate sufficient coverage of the tumor genome, because most of the reads would cover DNA from noncancerous cells. For WGS this additional amount of sequence coverage is currently too expensive for routine use. For CTCs, the problem is obtaining a sufficient number of pure tumor cells. In addition, some cancers appear to have no CTCs that can be isolated (Nagrath et al., 2007; Nakagawa et al., 2007; Riethdorf et al., 2007; Wendel et al., 2012; Rack et al., 2014).

Second, cancers can have millions of somatic mutations and these mutations are often not shared among all cells within a tumor or between metastatic sites. This can make it difficult to determine what the most important mutations are to target. One possible solution to this is whole-genome analysis of multiple minimally invasive biopsies (for advanced WGS such as LFR, only about 10 tumor cells are required) from a single tumor or in the case of late-stage cancers from multiple metastatic sites as well as the original tumor. Mutations that are shared among all sites are likely to be better targets for therapy because they are found in all cancer cells. In cases where it is not possible to obtain multiple biopsies, CTCs might be used. Studies have shown that at least five CTCs can be collected from approximately 50% of cancers of all stages (Nagrath et al., 2007; Nakagawa et al., 2007; Riethdorf et al., 2007; Wendel et al., 2012; Rack et al., 2014). As has been discussed earlier in this chapter, starting from a small number of cells can result in many false-positive variants. Nevertheless methods such as LFR can be used to remove many of these false-positives and identify the true somatic variants. This was demonstrated in our study with LFR using as few as five CTCs from a patient with breast cancer (Gulbahce et al., paper under review). A completely personalized combination therapy was determined for a patient's cancer targeting both driver and passenger mutations found to be present in all CTCs. Such targeted combination therapies, potentially including personalized cancer

vaccines, that are effective against all cancer lineages provide the best way to overpower the tumor in the first treatment before resistance forms. This type of comprehensive personalized analysis and therapy strategy would not have been possible with current panel sequencing approaches.

9.2.4 Advanced Whole-Genome Sequencing as Ultimate Genetic Test for Health Preservation

Our genomes contain all of the genetic instructions for development and functioning of all our tissues, and thus affect all aspects of our lives. They comprise the complete set of genetic instructions inherited from our parents in a "genetic lottery." For this reason, WGS enables the most informative interpretation of our genetic makeup and results in the most precise interventions needed to counterbalance some of our weak genetic spots. As such we have to start using this "ultimate" genetic test (Drmanac, 2012) as early as practical in our lives.

High-quality WGS is already available and has become affordable for many people. In the near future we will have "perfect" (practically 100% accurate and complete) personal WGS (Peters et al., 2014) with more precise interpretation and at an even more affordable price. The emerging trend in health care is precision medicine informed by personal genomic information. In this approach all decisions related to health care, from preventive actions to diagnostic schedules and treatment opinions, depend on the instructions encoded in our genome. As a result, it is becoming critically important for each of us to have our genome sequenced before our fifties when we start using more prescription drugs, or before our thirties when the incidence of cancer starts increasing, or before our twenties to influence our reproductive choices, or at birth to prevent some early childhood diseases, or prenatally (e.g., in IVF) to select our healthiest genome. Essentially the sooner individuals have their genome sequenced, the more health and other benefits they can derive.

Once individuals have their genome sequenced and securely stored, they can start using it to promote a longer, healthier, and more productive and satisfactory life. In the most advanced implementation of personal genomics, automatic actionable reports will be generated when needed, avoiding overinterpretation and the resulting unnecessary and costly interventions or anxieties. For example, if the genome of a newborn is sequenced and she is determined to be a carrier of a disease, that information would not be reported to the parent or child until immediately before reproductive age because there are no actions that can be taken earlier. The opposite would be true for cases in which mutations in the genome can cause congenital heart defects or early-onset sudden heart failure (Nyegaard et al., 2012; Al Turki et al., 2014), early cancers (Shah et al., 2013; Zhang et al., 2015), autism (O'Roak et al., 2011, 2012; Sanders et al., 2012; Michaelson et al., 2012; Yang et al., 2013; Iossifov et al., 2014; Al Turki et al., 2014; De Rubeis et al., 2014), intolerance of anesthesia (Rosenberg et al., 2015), and many others (Ciszkowski et al., 2009; Worthey et al., 2011; Epi et al., 2013; Gilissen et al., 2014). In these cases reports suggesting measures that can be taken or warning against taking others will be generated immediately for the parents, but will also be available to medical doctors and caregivers whenever needed. It is becoming more evident that such presymptomatic reports can help prevent disease development. Developing validated and approved genome interpretation software providing dependable quality is critical for the successful implementation of precision medicine.

Another important aspect of individuals having their genome sequenced as early as practical is to maximize preventive health care with minimal costs. A genome sequence enables targeted medical follow-up to people who need them. For example, *BRCA1* and *BRCA2* mutations are highly predictive of early-onset breast cancer and occur in 1 in 500 women. Because this is a relatively low frequency in the general population, it is too costly to discuss genetic mutations in these genes and possible courses of action with every women. Instead, only women with a family history of breast cancer are given genetic counseling, but this criterion addresses only about 50% of women with cancer predisposing *BRCA1* and *BRCA2* mutations. However, if by default every girl had her genome sequenced, every individual harboring these cancer predisposing mutations would be selectively referred to counseling and needed interventions. Those who did not have the mutations, including those with a family history, would not receive counseling because it would be unnecessary. It is becoming evident that more cancers affecting both women and men have inherited genetic components (Karageorgos et al., 2015). Importantly, the WGS test provides all genetic information not just for cancer or a few specific diseases, but for all diseases.

The genome program is an open, signal-processing program that allows for adaptive functioning of all our tissues. For this reason, the state of our tissues in any moment depends on our "past life" as processed by our genome program. Furthermore, somatic errors (genetic or epigenetic mutations) or infections occur in a stochastic manner as modulated by environmental factors. Many of these changes are irreversible or almost irreversible, resulting in genetic or epigenetic "lock-in" of cell lineages. Impactful changes of this nature in a tissue can cause a cascade of changes in other tissues as a result of the co-dependent nature of our organs and tissues. As a result of all of these interdependences, genetic predictions are conditional. To have a more precise assessment of our health

we need to measure actual molecular states of our tissues regularly: for example, the transcriptome or the epigenome of white blood cells or the micro-/mycobiome of the gut (Chen et al., 2012). These individual "omics" tests (iOmics) are as important as having a WGS test. The advanced iOmics tests can be performed on many tissues, including separate analyses of thousands of single blood cells or rare circulating cells (e.g., CTCs) or deep sequencing and epigenetic analysis of cfDNA. Combining WGS with iOmics tests provides the most informative prognostics and diagnostics (Drmanac, 2012).

9.3 WHAT IS NEEDED TO IMPLEMENT THIS VISION OF GENOMIC PRECISION HEALTH CARE FULLY?

9.3.1 Perfecting and Scaling Personal Whole-Genome Sequencing

As mentioned previously in this chapter, we are still far from achieving a fully accurate and complete personal genome. Less than "perfect" genomes can be used for precision health care but many variants will be missing and many false-positive variants will be present. These types of errors will affect variant interdependence analyses and necessitate the validation of all critical variants to ensure they are real. This adds unnecessary cost and complexity to the process. As discussed earlier, advanced LFR with over 10,000 barcodes and a phased de novo genome assembly is needed to achieve "perfect" WGS inexpensively (Peters et al., 2014). Highly efficient methods have been proposed to use 1 million or more barcodes in a single tube (Drmanac et al., 2013). This would allow the scaling of LFR to millions of "perfect" personal genomes per year. In addition to improving and advancing co-barcoding processes, more efficient sequencers than are currently available are necessary to generate millions of "perfect" genomes requiring at least 100× read coverage. This scaling can be achieved through the same DNB nanoarrays described earlier. Combined with cost-effective large-scale manufacturing by BGI (Complete Genomics was acquired by BGI in 2013) in China, this technology is poised to provide cost-effective instruments and sequencing reagents for the broad worldwide use of advanced genetic tests and personal WGS.

Another important component for scaled WGS is the integrated infrastructure from samples to reports including data management and automated interpretation. We suggest that every country invest in the large-scale personal WGS and iOmics infrastructure much as countries invested in the communication and broadband Internet infrastructure. This is critical for future precision health care and the growth of many other industries (e.g., food and biotechnology).

9.3.2 Building a Gigantic Knowledge Base and Genome "Compiler"

The same population-scale sequencing infrastructure will enable the accumulation of data and knowledge needed to perform more accurate and complete interpretation of personal genome and iOmics tests. This will require a gigantic database composed of WGS and iOmics tests as well as health and other phenotype and environmental measurements repeatedly obtained for millions of individuals over their lifetime. This is necessary to be able to observe molecular changes from healthy to disease states. Projects of this nature have been initiated by research institutions such as Lee Hood's Institute for Systems Biology in Seattle (Hood and Price, 2014). "Big Data" such as this will allow the use of machine learning and artificial intelligence methods to reverse engineer our human genome program.

As proposed by one of us (R.D.), the ultimate long-term result of such reverse engineering and decoding efforts would be a human genome "compiler" that would be able to create a computer-"executable" individual genome program from an individual WGS. "Compiling" personal genomes can be used in various ways as described subsequently. To speed these efforts, the formation of a partnership among millions of people, data generation companies, and data management companies is necessary. We propose that individuals pay a minimal price for their WGS/iOmics tests, sufficient to cover the actual cost, and allow their data to be aggregated to improve the reports of everyone in the database. To attract participation it can be arranged for early adopters to pay less than future participants for advanced interpretation of their WGS/iOmics tests.

Having a comprehensive understanding of the genetic and molecular pathways of tissue development, aging, and disease initiation and progression will provide a rational basis for designing precise personal disease prevention and treatment programs (from diet and supplements to drugs, psychotherapy, or cell and gene therapies) that take into account peculiarities of our genome and molecular conditions of our tissues.

9.3.3 Running Our Genome Program In Silico and In Vitro

Once a human "genome compiler" is developed, we can use it to examine the "performance" of a personal genome (Fig. 9.7). A personal genome would be compiled to generate an executable program that would be run in silico to observe the molecular and physiologic phenotypes defined by that genome under specified environmental conditions. Such observations in principal can provide much more accurate "interpretation of personal WGS" than the correlation analysis performed currently. Running our

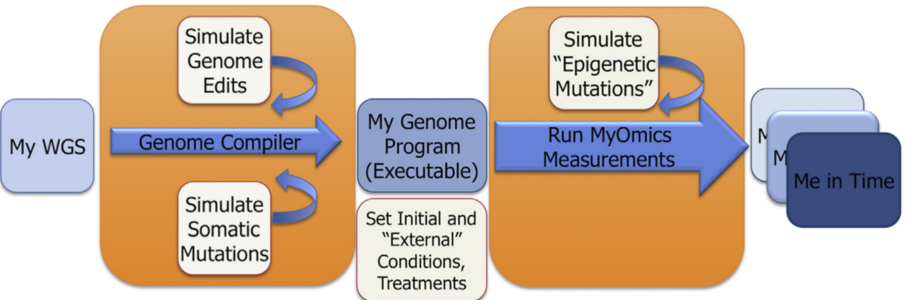


FIGURE 9.7 Future genomic prognostics. A personal genome sequence directly or with some test edits is compiled in an executable computer program. After setting initial conditions (e.g., epigenetic and cytoplasmic states created by merging egg and sperm as well as external environmental conditions including potential treatments), the executable program is run in silico to observe changes in the molecular states of simulated tissues over time from development to aging. This approach of observing actual outcomes is expected to provide contextual personal genome interpretation that is more accurate than correlation analysis. The iterative process of using current states of the genome, tissues, and environment as inputs to compute a new tissue/organism state allows: (1) adjusting simulation predictions by actual measurements of various "omics" tests or (2) simulating epigenetic "mutations." *WGS*, whole-genome sequencing.

genome program allows all pathway interactions to be taken into account. The "genome compiler" allows simulation experiments by editing our genomes in silico at certain positions, or therapy screens in silico. The full contextual impact of one or multiple genome changes can be observed. Using the clustered regularly interspaced short palindromic repeats system (Cong et al., 2013), genome edits with desirable effects can potentially be introduced into our induced pluripotent stem cells to correct for some genetic deficiencies. Such edits can be further evaluated in vitro in the rapidly developing personal "organs-on-chips" technology (Bhatia and Ingber, 2014).

9.3.4 Virtual Health Coaching

To be successful, in addition to defining precise personalized preventive actions or treatments, the delivery and implementation of genomic precision medicine must be personalized to accommodate various psychological or cultural preferences and beliefs. Furthermore, continuous and customized education of the users is a critical aspect of this implementation. This level of personalized and customized delivery of medical information, which requires the strong participation and education of users, would be too expensive if performed by traditional health care providers and would be available only to a limited number of people. To make adaptable and affordable precision medicine support available to millions of people, we propose a virtual health coaching sponsor and assistant as an inexpensive Web service available any time of day on personal devices with maximal privacy (Drmanac and Drmanac, 2009). We hypothesize that procedural aspects of the traditional coaching process can be efficiently implemented for genomic medicine education and health coaching through a personalized virtual sponsor using today's computational capabilities. The health-coaching sponsor is designed to provide services and information to people when they need it and in a form that works for different personalities and preferences (Fig. 9.8).

The proposed virtual coach system is composed of coaching sessions with sets of important mind-opening questions or statements that help users to take health into their own hands. This includes developing plans with a focus on disease prevention through genetic testing and the introduction of healthy habits from proper diet and exercise (Jones et al., 2016) to needed relaxation and sleep. The virtual sponsor also serves as a health assistant and provides users with important reminders related to health, diet, supplements, exercise, and relaxation, and schedules doctor's visits or checkups in accordance with an individual's health needs and health plan. Furthermore, coaching sessions and guided processes are intertwined with access to critical functional knowledge and information.

It is our belief that virtual health coaching will preserve and enhance all proven benefits of professional coaching while removing coach biases and other limitations. At the same time, this service can be affordable to millions of people. Two critical features of virtual coaching are customization and personalization of the virtual coach and coaching sessions. Users will be able to select (customize) coach attributes according to their preferences, such as voice, tempo, language, level of details, or even facial features. Furthermore, the coaching program will automatically adapt according to detected user preferences to provide a more effective personalized coaching. Such levels of customization and personalization will facilitate the adoption of coaching as a critical enabler of personalized and participatory precision health care.

The capabilities of a virtual sponsor approach could be used in various areas of genomic health to put genetic and genomic information into the life context of individuals and their families. In addition to education in medical genomics and personalized medicine, the proposed virtual sponsor has the potential to become an assistant to genetic

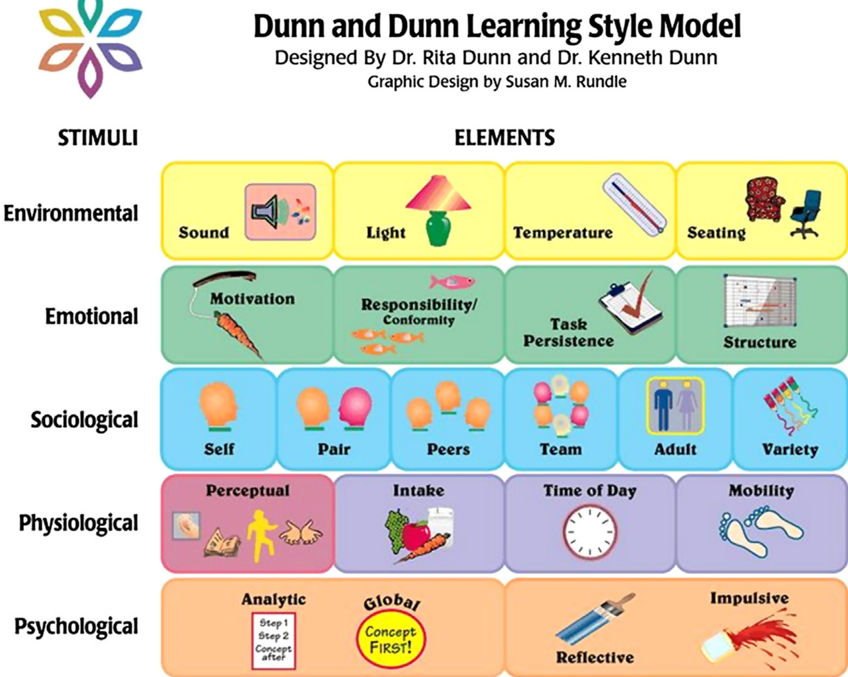


FIGURE 9.8 Dunn and Dunn preference model. The five stimuli groups or dimensions encompass environmental, emotional, sociological, physiological, and psychological areas. The stimuli deal with how people perceive, interact, and respond within their environment. Within those groups there are 21 variables or elements for which a person as a learner and user may have various levels of preference.

counselors and provide coaching for healthy habits and attitudes. Coaching, education, genomic tests, medical family history, and health information resources could easily be accessible via virtual sponsor software to empower people to make informed health decisions and satisfy specific health needs. Thus, virtual sponsoring services can reduce unsustainable health care costs of individualized medicine, educate consumers about appropriate testing and public health services, and help individuals navigate the complex health care delivery system.

9.4 CONCLUSION

In large part, the genetic program encoded in our genomes defines our lives, from health and body strength to mental abilities and behaviors. It is now possible to read this program by WGS with increasing efficiency, scale, completeness, and accuracy even from just a few cells. Already, important prognostic and diagnostic values of personal WGS and other iOmics can be seen. The ability to interpret these tests is also growing exponentially as genetic databases increase in size and experimental methods improve. Furthermore, we can efficiently edit our genomes with single-base precision. Human societies are poised to use these advances in genomics and molecular biology as well as the power of virtual coaches and assistants to implement a broad and cost-effective genomic precision health care focused on personalized disease prevention, from reducing the prevalence of genetic diseases in newborns to extending our healthy life span. What is needed is a joint effort by governments, research and medical organizations, industry, and consumers for expedient and balanced (i.e., sacrificing some sensitivity for specificity) implementation of this advanced health care. Already there have been several encouraging efforts in the United Kingdom, United States, China, The Netherlands, and other countries.

REFERENCES

Ahn, S.M., Kim, T.H., Lee, S., Kim, D., Ghang, H., Kim, D.S., Kim, B.C., Kim, S.Y., Kim, W.Y., Kim, C., Park, D., Lee, Y.S., Kim, S., Reja, R., Jho, S., Kim, C.G., Cha, J.Y., Kim, K.H., Lee, B., Bhak, J., Kim, S.J., 2009. The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group. Genome Res. 19, 1622−1629.

Al Turki, S., Manickaraj, A.K., Mercer, C.L., Gerety, S.S., Hitz, M.P., Lindsay, S., D'Alessandro, L.C., Swaminathan, G.J., Bentham, J., Arndt, A.K., Low, J., Breckpot, J., Gewillig, M., Thienpont, B., Abdul-Khaliq, H., Harnack, C., Hoff, K., Kramer, H.H., Schubert, S., Siebert, R., Toka, O., Cosgrove, C., Watkins, H., Lucassen, A.M., O'Kelly, I.M., Salmon, A.P., BU'lock, F.A., Granados-Riveron, J., Setchfield, K., Thornborough, C., Brook, J.D., Mulder, B., Klaassen, S., Bhattacharya, S., Devriendt, K., Fitzpatrick, D.F., Consortium, U. K, Wilson, D.I., Mital, S., Hurles, M.E., 2014. Rare variants in *NR2F2* cause congenital heart defects in humans. Am. J. Hum. Genet. 94, 574−585.

Amini, S., Pushkarev, D., Christiansen, L., Kostem, E., Royce, T., Turk, C., Pignatelli, N., Adey, A., Kitzman, J.O., Vijayan, K.,

Ronaghi, M., Shendure, J., Gunderson, K.L., Steemers, F.J., 2014. Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing. Nat. Genet. 46, 1343–1349.

Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., Boutell, J.M., Bryant, J., Carter, R.J., Keira Cheetham, R., Cox, A.J., Ellis, D.J., Flatbush, M.R., Gormley, N.A., Humphray, S.J., Irving, L.J., Karbelashvili, M.S., Kirk, S.M., Li, H., Liu, X., Maisinger, K.S., Murray, L.J., Obradovic, B., Ost, T., Parkinson, M.L., Pratt, M.R., Rasolonjatovo, I.M., Reed, M.T., Rigatti, R., Rodighiero, C., Ross, M.T., Sabot, A., Sankar, S.V., Scally, A., Schroth, G.P., Smith, M.E., Smith, V.P., Spiridou, A., Torrance, P.E., Tzonev, S.S., Vermaas, E.H., Walter, K., Wu, X., Zhang, L., Alam, M.D., Anastasi, C., Aniebo, I.C., Bailey, D.M., Bancarz, I.R., Banerjee, S., Barbour, S.G., Baybayan, P.A., Benoit, V.A., Benson, K.F., Bevis, C., Black, P.J., Boodhun, A., Brennan, J.S., Bridgham, J.A., Brown, R.C., Brown, A.A., Buermann, D.H., Bundu, A.A., Burrows, J.C., Carter, N.P., Castillo, N., Chiara, E.C.M., Chang, S., Neil Cooley, R., Crake, N.R., Dada, O.O., Diakoumakos, K.D., Dominguez-Fernandez, B., Earnshaw, D.J., Egbujor, U.C., Elmore, D.W., Etchin, S.S., Ewan, M.R., Fedurco, M., Fraser, L.J., Fuentes Fajardo, K.V., Scott Furey, W., George, D., Gietzen, K.J., Goddard, C.P., Golda, G.S., Granieri, P.A., Green, D.E., Gustafson, D.L., Hansen, N.F., Harnish, K., Haudenschild, C.D., Heyer, N.I., Hims, M.M., Ho, J.T., Horgan, A.M., et al., 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456, 53–59.

Bhatia, S.N., Ingber, D.E., 2014. Microfluidic organs-on-chips. Nat. Biotechnol. 32, 760–772.

Carnevali, P., Baccash, J., Halpern, A.L., Nazarenko, I., Nilsen, G.B., Pant, K.P., Ebert, J.C., Brownley, A., Morenzoni, M., Karpinchyk, V., Martin, B., Ballinger, D.G., Drmanac, R., 2012. Computational techniques for human genome resequencing using mated gapped reads. J. Comput. Biol. 19, 279–292.

Centers for disease control and prevention, 2011. A. S. F. R. M., Society for Assisted Reproductive Technology. U.S. Department of Health and Human Services, 2009 Assisted Reproductive Technology Success Rates: National Summary and Fertility Clinic Reports.

Chen, R., Mias, G.I., LI-Pook-Than, J., Jiang, L., Lam, H.Y., Chen, R., Miriami, E., Karczewski, K.J., Hariharan, M., Dewey, F.E., Cheng, Y., Clark, M.J., Im, H., Habegger, L., Balasubramanian, S., O'huallachain, M., Dudley, J.T., Hillenmeyer, S., Haraksingh, R., Sharon, D., Euskirchen, G., Lacroute, P., Bettinger, K., Boyle, A.P., Kasowski, M., Grubert, F., Seki, S., Garcia, M., Whirl-Carrillo, M., Gallardo, M., Blasco, M.A., Greenberg, P.L., Snyder, P., Klein, T.E., Altman, R.B., Butte, A.J., Ashley, E.A., Gerstein, M., Nadeau, K.C., Tang, H., Snyder, M., 2012. Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell 148, 1293–1307.

Cheung, S.W., Patel, A., Leung, T.Y., 2015. Accurate description of DNA-based noninvasive prenatal screening. N. Engl. J. Med. 372, 1675–1677.

Christianson, A., Howson, C.P., Modell, B., 2006. The Hidden Toll of Dying and Disabled Children. March of Dimes Global Report on Birth Defects (March of Dimes Birth Defects Foundation).

Ciszkowski, C., Madadi, P., Phillips, M.S., Lauwers, A.E., Koren, G., 2009. Codeine, ultrarapid-metabolism genotype, and postoperative death. N. Engl. J. Med. 361, 827–828.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., Zhang, F., 2013. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823.

Crow, J.F., 2000. The origins, patterns and implications of human spontaneous mutation. Nat. Rev. Genet. 1, 40–47.

De Ligt, J., Willemsen, M.H., Van Bon, B.W., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-Van Silfhout, A.T., Koolen, D.A., De Vries, P., Gilissen, C., Del Rosario, M., Hoischen, A., Scheffer, H., De Vries, B.B., Brunner, H.G., Veltman, J.A., Vissers, L.E., 2012. Diagnostic exome sequencing in persons with severe intellectual disability. N. Engl. J. Med. 367, 1921–1929.

De Mouzon, J., Goossens, V., Bhattacharya, S., Castilla, J.A., Ferraretti, A.P., Korsak, V., Kupka, M., Nygren, K.G., Andersen, A.N., 2012. Assisted reproductive technology in Europe, 2007: results generated from European registers by ESHRE. Hum. Reprod. 27, 954–966.

De Mouzon, J., Lancaster, P., Nygren, K.G., Sullivan, E., Zegers-Hochschild, F., Mansour, R., Ishihara, O., Adamson, D., 2009. World collaborative report on assisted reproductive technology, 2002. Hum. Reprod. 24, 2310–2320.

De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Ercument Cicek, A., Kou, Y., Liu, L., Fromer, M., Walker, S., Singh, T., Klei, L., Kosmicki, J., Fu, S.C., Aleksic, B., Biscaldi, M., Bolton, P.F., Brownfeld, J.M., Cai, J., Campbell, N.G., Carracedo, A., Chahrour, M.H., Chiocchetti, A.G., Coon, H., Crawford, E.L., Crooks, L., Curran, S.R., Dawson, G., Duketis, E., Fernandez, B.A., Gallagher, L., Geller, E., Guter, S.J., Sean Hill, R., Ionita-Laza, I., Jimenez Gonzalez, P., Kilpinen, H., Klauck, S.M., Kolevzon, A., Lee, I., Lei, J., Lehtimaki, T., Lin, C.F., MA'ayan, A., Marshall, C.R., Mcinnes, A.L., Neale, B., Owen, M.J., Ozaki, N., Parellada, M., Parr, J.R., Purcell, S., Puura, K., Rajagopalan, D., Rehnstrom, K., Reichenberg, A., Sabo, A., Sachse, M., Sanders, S.J., Schafer, C., Schulte-Ruther, M., Skuse, D., Stevens, C., Szatmari, P., Tammimies, K., Valladares, O., Voran, A., Wang, L.S., Weiss, L.A., Jeremy Willsey, A., Yu, T.W., Yuen, R.K., DDD Study, Homozygosity Mapping Collaborative For Autism, UK Consortium, The Autism Sequencing C, Cook, E.H., Freitag, C.M., Gill, M., Hultman, C.M., Lehner, T., Palotie, A., Schellenberg, G.D., Sklar, P., State, M.W., Sutcliffe, J.S., Walsh, C.A., Scherer, S.W., Zwick, M.E., Barrett, J.C., Cutler, D.J., Roeder, K., Devlin, B., Daly, M.J., Buxbaum, J.D., 2014. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature 209–215.

Drmanac, R., 2006. Nucleic Acid Analysis by Random Mixtures of Non-Overlapping Fragments. WO 2006/138284 A2.

Drmanac, R., 2012. Medicine. The ultimate genetic test. Science 336, 1110–1112.

Drmanac, R., Callow, M.J., Drmanac, S., 2007. High Throughput Genome Sequencing on DNA Arrays. WO/2007/133831.

Drmanac, R., Crkvenjakov, R., 1990. Prospects for a miniaturized, simplified, and frugal human genome project. Sci. Yugosl. 16, 97–107.

Drmanac, S., Drmanac, R., 2009. Systems, Computer Readable Program Products, and Computer Implemented Methods to Facilitate On-Demand, User-driven, Virtual Sponsoring Sessions for One or More User Selected Topics Through User Designed Virtual Sponsors. United States patent application. US 13/384,090.

Drmanac, R., Peters, B.A., Alexeev, A., 2013. Multiple Tagging of Individual Long DNA Fragments. US20140323316 A1.

Drmanac, R., Peters, B.A., Alexeev, A., Hong, P., 2010a. Methods and Compositions for Long Fragment Read Sequencing. US 8592150 B2.

Drmanac, R., Sparks, A.B., Callow, M.J., Halpern, A.L., Burns, N.L., Kermani, B.G., Carnevali, P., Nazarenko, I., Nilsen, G.B., Yeung, G., Dahl, F., Fernandez, A., Staker, B., Pant, K.P., Baccash, J., Borcherding, A.P., Brownley, A., Cedeno, R., Chen, L., Chernikoff, D., Cheung, A., Chirita, R., Curson, B., Ebert, J.C., Hacker, C.R., Hartlage, R., Hauser, B., Huang, S., Jiang, Y., Karpinchyk, V., Koenig, M., Kong, C., Landers, T., Le, C., Liu, J., Mcbride, C.E., Morenzoni, M., Morey, R.E., Mutch, K., Perazich, H., Perry, K., Peters, B.A., Peterson, J., Pethiyagoda, C.L., Pothuraju, K., Richter, C., Rosenbaum, A.M., Roy, S., Shafto, J., Sharanhovich, U., Shannon, K.W., Sheppy, C.G., Sun, M., Thakuria, J.V., Tran, A., Vu, D., Zaranek, A.W., Wu, X., Drmanac, S., Oliphant, A.R., Banyai, W.C., Martin, B., Ballinger, D.G., Church, G.M., Reid, C.A., 2010b. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science 327, 78–81.

Epi, K.C., Epilepsy Phenome/Genome Project, Allen, A.S., Berkovic, S.F., Cossette, P., Delanty, N., Dlugos, D., Eichler, E.E., Epstein, M.P., Glauser, T., Goldstein, D.B., Han, Y., Heinzen, E.L., Hitomi, Y., Howell, K.B., Johnson, M.R., Kuzniecky, R., Lowenstein, D.H., Lu, Y.F., Madou, M.R., Marson, A.G., Mefford, H.C., Esmaeeli Nieh, S., O'Brien, T.J., Ottman, R., Petrovski, S., Poduri, A., Ruzzo, E.K., Scheffer, I.E., Sherr, E.H., Yuskaitis, C.J., Abou-Khalil, B., Alldredge, B.K., Bautista, J.F., Berkovic, S.F., Boro, A., Cascino, G.D., Consalvo, D., Crumrine, P., Devinsky, O., Dlugos, D., Epstein, M.P., Fiol, M., Fountain, N.B., French, J., Friedman, D., Geller, E.B., Glauser, T., Glynn, S., Haut, S.R., Hayward, J., Helmers, S.L., Joshi, S., Kanner, A., Kirsch, H.E., Knowlton, R.C., Kossoff, E.H., Kuperman, R., Kuzniecky, R., Lowenstein, D.H., Mcguire, S.M., Motika, P.V., Novotny, E.J., Ottman, R., Paolicchi, J.M., Parent, J.M., Park, K., Poduri, A., Scheffer, I.E., Shellhaas, R.A., Sherr, E.H., Shih, J.J., Singh, R., Sirven, J., Smith, M.C., Sullivan, J., Lin Thio, L., Venkat, A., Vining, E.P., Von Allmen, G.K., Weisenberg, J.L., Widdess-Walsh, P., Winawer, M.R., 2013. De novo mutations in epileptic encephalopathies. Nature 501, 217–221.

ESHRE, 2012. Assisted Reproductive Technology and Intrauterine Inseminations in Europe, 2009: Results Generated From European Registers by ESHRE, Presented at 2012 Annual Meeting of Eshre, Istanbul.

Francioli, L.C., Polak, P.P., Koren, A., Menelaou, A., Chun, S., Renkens, I., Genome of The Netherlands, Consortium, Van Duijn, C.M., Swertz, M., Wijmenga, C., Van Ommen, G., Slagboom, P.E., Boomsma, D.I., Ye, K., Guryev, V., Arndt, P.F., Kloosterman, W.P., De Bakker, P.I., Sunyaev, S.R., 2015. Genome-wide patterns and properties of *de novo* mutations in humans. Nat. Genet. 47, 822–826.

Gilissen, C., Hehir-Kwa, J.Y., Thung, D.T., Van De Vorst, M., Van Bon, B.W., Willemsen, M.H., Kwint, M., Janssen, I.M., Hoischen, A., Schenck, A., Leach, R., Klein, R., Tearle, R., Bo, T., Pfundt, R., Yntema, H.G., De Vries, B.B., Kleefstra, T., Brunner, H.G., Vissers, L.E., Veltman, J.A., 2014. Genome sequencing identifies major causes of severe intellectual disability. Nature 511, 344–347.

Gratten, J., Wray, N.R., Keller, M.C., Visscher, P.M., 2014. Large-scale genomics unveils the genetic architecture of psychiatric disorders. Nat. Neurosci. 17, 782–790.

Hamilton, B.A., Yu, B.D., 2012. Modifier genes and the plasticity of genetic networks in mice. PLoS Genet. 8, e1002644.

Hassold, T., Hunt, P., 2009. Maternal age and chromosomally abnormal pregnancies: what we know and what we wish we knew. Curr. Opin. Pediatr. 21, 703–708.

Hood, L., Price, N.D., 2014. Demystifying disease, democratizing health care. Sci. Transl. Med. 6, 225ed5.

Iossifov, I., O'Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E., Smith, J.D., Paeper, B., Nickerson, D.A., Dea, J., Dong, S., Gonzalez, L.E., Mandell, J.D., Mane, S.M., Murtha, M.T., Sullivan, C.A., Walker, M.F., Waqar, Z., Wei, L., Willsey, A.J., Yamrom, B., Lee, Y.H., Grabowska, E., Dalkic, E., Wang, Z., Marks, S., Andrews, P., Leotta, A., Kendall, J., Hakker, I., Rosenbaum, J., Ma, B., Rodgers, L., Troge, J., Narzisi, G., Yoon, S., Schatz, M.C., Ye, K., Mccombie, W.R., Shendure, J., Eichler, E.E., State, M.W., Wigler, M., 2014. The contribution of *de novo* coding mutations to autism spectrum disorder. Nature 515, 216–221.

Jiang, Y.H., Yuen, R.K., Jin, X., Wang, M., Chen, N., Wu, X., Ju, J., Mei, J., Shi, Y., He, M., Wang, G., Liang, J., Wang, Z., Cao, D., Carter, M.T., Chrysler, C., Drmic, I.E., Howe, J.L., Lau, L., Marshall, C.R., Merico, D., Nalpathamkalam, T., Thiruvahindrapuram, B., Thompson, A., Uddin, M., Walker, S., Luo, J., Anagnostou, E., Zwaigenbaum, L., Ring, R.H., Wang, J., Lajonchere, C., Wang, J., Shih, A., Szatmari, P., Yang, H., Dawson, G., Li, Y., Scherer, S.W., 2013. Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. Am. J. Hum. Genet. 93, 249–263.

Jones, N., Kiely, J., Suraci, B., Collins, D.J., De Lorenzo, D., Pickering, C., Grimaldi, K.A., 2016. A genetic-based algorithm for personalized resistance training. Biol. Sport 33, 117–126.

Kaper, F., Swamy, S., Klotzle, B., Munchel, S., Cottrell, J., Bibikova, M., Chuang, H.Y., Kruglyak, S., Ronaghi, M., Eberle, M.A., Fan, J.B., 2013. Whole-genome haplotyping by dilution, amplification, and sequencing. Proc. Natl. Acad. Sci. U. S. A. 110, 5552–5557.

Karageorgos, I., Mizzi, C., Giannopoulou, E., Pavlidis, C., Peters, B.A., Zagoriti, Z., Stenson, P.D., Mitropoulos, K., Borg, J., Kalofonos, H.P., Drmanac, R., Stubbs, A., van der Spek, P., Cooper, D.N., Katsila, T., Patrinos, G.P., 2015. Identification of cancer predisposition variants in apparently healthy individuals using a next-generation sequencing-based family genomics approach. Hum. Genomics 9, 12.

Kim, J.I., Ju, Y.S., Park, H., Kim, S., Lee, S., Yi, J.H., Mudge, J., Miller, N.A., Hong, D., Bell, C.J., Kim, H.S., Chung, I.S., Lee, W.C., Lee, J.S., Seo, S.H., Yun, J.Y., Woo, H.N., Lee, H., Suh, D., Kim, H.J., Yavartanoo, M., Kwak, M., Zheng, Y., Lee, M.K., Kim, J.Y., Gokcumen, O., Mills, R.E., Zaranek, A.W., Thakuria, J., Wu, X., Kim, R.W., Huntley, J.J., Luo, S., Schroth, G.P., Wu, T.D., Kim, H., Yang, K.S., Park, W.Y., Church, G.M., Lee, C., Kingsmore, S.F., Seo, J.S., 2009. A highly annotated whole-genome sequence of a Korean individual. Nature 460, 1011–1015.

Kong, A., Frigge, M.L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S.A., Sigurdsson, A., Jonasdottir, A., Wong, W.S., Sigurdsson, G., Walters, G.B., Steinberg, S., Helgason, H., Thorleifsson, G., Gudbjartsson, D.F., Helgason, A., Magnusson, O.T., Thorsteinsdottir, U., Stefansson, K., 2012. Rate of *de novo* mutations and the importance of father's age to disease risk. Nature 488, 471–475.

Kuleshov, V., Xie, D., Chen, R., Pushkarev, D., Ma, Z., Blauwkamp, T., Kertesz, M., Snyder, M., 2014. Whole-genome haplotyping using long reads and statistical methods. Nat. Biotechnol. 32, 261−266.

Levene, M.J., Korlach, J., Turner, S.W., Foquet, M., Craighead, H.G., Webb, W.W., 2003. Zero-mode waveguides for single-molecule analysis at high concentrations. Science 299, 682−686.

Ley, T.J., Mardis, E.R., Ding, L., Fulton, B., Mclellan, M.D., Chen, K., Dooling, D., Dunford-Shore, B.H., Mcgrath, S., Hickenbotham, M., Cook, L., Abbott, R., Larson, D.E., Koboldt, D.C., Pohl, C., Smith, S., Hawkins, A., Abbott, S., Locke, D., Hillier, L.W., Miner, T., Fulton, L., Magrini, V., Wylie, T., Glasscock, J., Conyers, J., Sander, N., Shi, X., Osborne, J.R., Minx, P., Gordon, D., Chinwalla, A., Zhao, Y., Ries, R.E., Payton, J.E., Westervelt, P., Tomasson, M.H., Watson, M., Baty, J., Ivanovich, J., Heath, S., Shannon, W.D., Nagarajan, R., Walter, M.J., Link, D.C., Graubert, T.A., Dipersio, J.F., Wilson, R.K., 2008. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature 456, 66−72.

Mak, A.C., Lai, Y.Y., Lam, E.T., Kwok, T.P., Leung, A.K., Poon, A., Mostovoy, Y., Hastie, A.R., Stedman, W., Anantharaman, T., Andrews, W., Zhou, X., Pang, A.W., Dai, H., Chu, C., Lin, C., Wu, J.J., Li, C.M., Li, J.W., Yim, A.K., Chan, S., Sibert, J., Dzakula, Z., Cao, H., Yiu, S.M., Chan, T.F., Yip, K.Y., Xiao, M., Kwok, P.Y., 2016. Genome-wide structural variation detection by genome mapping on nanochannel arrays. Genetics 202, 351−362.

Matthews, T.J., Macdorman, M.F., Thoma, M.E., 2015. Infant mortality statistics from the 2013 period linked birth/infant death data set. Natl. Vital Stat. Rep. 64, 1−30.

Mckernan, K.J., Peckham, H.E., Costa, G.L., Mclaughlin, S.F., Fu, Y., Tsung, E.F., Clouser, C.R., Duncan, C., Ichikawa, J.K., Lee, C.C., Zhang, Z., Ranade, S.S., Dimalanta, E.T., Hyland, F.C., Sokolsky, T.D., Zhang, L., Sheridan, A., Fu, H., Hendrickson, C.L., Li, B., Kotler, L., Stuart, J.R., Malek, J.A., Manning, J.M., Antipova, A.A., Perez, D.S., Moore, M.P., Hayashibara, K.C., Lyons, M.R., Beaudoin, R.E., Coleman, B.E., Laptewicz, M.W., Sannicandro, A.E., Rhodes, M.D., Gottimukkala, R.K., Yang, S., Bafna, V., Bashir, A., Macbride, A., Alkan, C., Kidd, J.M., Eichler, E.E., Reese, M.G., De La Vega, F.M., Blanchard, A.P., 2009. Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. Genome Res. 19, 1527−1541.

Michaelson, J.J., Shi, Y., Gujral, M., Zheng, H., Malhotra, D., Jin, X., Jian, M., Liu, G., Greer, D., Bhandari, A., Wu, W., Corominas, R., Peoples, A., Koren, A., Gore, A., Kang, S., Lin, G.N., Estabillo, J., Gadomski, T., Singh, B., Zhang, K., Akshoomoff, N., Corsello, C., Mccarroll, S., Iakoucheva, L.M., Li, Y., Wang, J., Sebat, J., 2012. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. Cell 151, 1431−1442.

Miller, N.A., Farrow, E.G., Gibson, M., Willig, L.K., Twist, G., Yoo, B., Marrs, T., Corder, S., Krivohlavek, L., Walter, A., Petrikin, J.E., Saunders, C.J., Thiffault, I., Soden, S.E., Smith, L.D., Dinwiddie, D.L., Herd, S., Cakici, J.A., Catreux, S., Ruehle, M., Kingsmore, S.F., 2015. A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. Genome Med. 7, 100.

Mizzi, C., Peters, B., Mitropoulou, C., Mitropoulos, K., Katsila, T., Agarwal, M.R., Van Schaik, R.H., Drmanac, R., Borg, J., Patrinos, G.P., 2014. Personalized pharmacogenomics profiling using whole-genome sequencing. Pharmacogenomics 15, 1223−1234.

Moore, G.E., 1965. Cramming more components onto integrated circuits. Electronics 38, 114−117.

Munne, S., Alikani, M., Tomkin, G., Grifo, J., Cohen, J., 1995. Embryo morphology, developmental rates, and maternal age are correlated with chromosome abnormalities. Fertil. Steril. 64, 382−391.

Nagrath, S., Sequist, L.V., Maheswaran, S., Bell, D.W., Irimia, D., Ulkus, L., Smith, M.R., Kwak, E.L., Digumarthy, S., Muzikansky, A., Ryan, P., Balis, U.J., Tompkins, R.G., Haber, D.A., Toner, M., 2007. Isolation of rare circulating tumour cells in cancer patients by microchip technology. Nature 450, 1235−1239.

Nakagawa, T., Martinez, S.R., Goto, Y., Koyanagi, K., Kitago, M., Shingai, T., Elashoff, D.A., Ye, X., Singer, F.R., Giuliano, A.E., Hoon, D.S., 2007. Detection of circulating tumor cells in early-stage breast cancer metastasis to axillary lymph nodes. Clin. Cancer Res. 13, 4105−4110.

Nyegaard, M., Overgaard, M.T., Sondergaard, M.T., Vranas, M., Behr, E.R., Hildebrandt, L.L., Lund, J., Hedley, P.L., Camm, A.J., Wettrell, G., Fosdal, I., Christiansen, M., Borglum, A.D., 2012. Mutations in calmodulin cause ventricular tachycardia and sudden cardiac death. Am. J. Hum. Genet. 91, 703−712.

O'Roak, B.J., Deriziotis, P., Lee, C., Vives, L., Schwartz, J.J., Girirajan, S., Karakoc, E., Mackenzie, A.P., Ng, S.B., Baker, C., Rieder, M.J., Nickerson, D.A., Bernier, R., Fisher, S.E., Shendure, J., Eichler, E.E., 2011. Exome sequencing in sporadic autism spectrum disorders identifies severe *de novo* mutations. Nat. Genet. 43, 585−589.

O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D., Turner, E.H., Stanaway, I.B., Vernot, B., Malig, M., Baker, C., Reilly, B., Akey, J.M., Borenstein, E., Rieder, M.J., Nickerson, D.A., Bernier, R., Shendure, J., Eichler, E.E., 2012. Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. Nature 485, 246−250.

Peters, B.A., Kermani, B.G., Alferov, O., Agarwal, M.R., Mcelwain, M.A., Gulbahce, N., Hayden, D.M., Tang, Y.T., Zhang, R.Y., Tearle, R., Crain, B., Prates, R., Berkeley, A., Munne, S., Drmanac, R., 2015. Detection and phasing of single base de novo mutations in biopsies from human in vitro fertilized embryos by advanced whole-genome sequencing. Genome Res. 25, 426−434.

Peters, B.A., Kermani, B.G., Sparks, A.B., Alferov, O., Hong, P., Alexeev, A., Jiang, Y., Dahl, F., Tang, Y.T., Haas, J., Robasky, K., Zaranek, A.W., Lee, J.H., Ball, M.P., Peterson, J.E., Perazich, H., Yeung, G., Liu, J., Chen, L., Kennemer, M.I., Pothuraju, K., Konvicka, K., Tsoupko-Sitnikov, M., Pant, K.P., Ebert, J.C., Nilsen, G.B., Baccash, J., Halpern, A.L., Church, G.M., Drmanac, R., 2012. Accurate whole-genome sequencing and haplotyping from 10 to 20 human cells. Nature 487, 190−195.

Peters, B.A., Liu, J., Drmanac, R., 2014. Co-barcoded sequence reads from long DNA fragments: a cost-effective solution for "perfect genome" sequencing. Front. Genet. 5, 466.

Pushkarev, D., Neff, N.F., Quake, S.R., 2009. Single-molecule sequencing of an individual human genome. Nat. Biotechnol. 27, 847−852.

Rack, B., Schindlbeck, C., Juckstock, J., Andergassen, U., Hepp, P., Zwingers, T., Friedl, T.W., Lorenz, R., Tesch, H., Fasching, P.A., Fehm, T., Schneeweiss, A., Lichtenegger, W., Beckmann, M.W., Friese, K., Pantel, K., Janni, W., Group, S.S., 2014. Circulating tumor

cells predict survival in early average-to-high risk breast cancer patients. J. Natl. Cancer Inst. 106.

Richards, J., Korgenski, E.K., Srivastava, R., Bonkowsky, J.L., 2015. Costs of the diagnostic odyssey in children with inherited leukodystrophies. Neurology 85, 1167–1170.

Riethdorf, S., Fritsche, H., Muller, V., Rau, T., Schindlbeck, C., Rack, B., Janni, W., Coith, C., Beck, K., Janicke, F., Jackson, S., Gornet, T., Cristofanilli, M., Pantel, K., 2007. Detection of circulating tumor cells in peripheral blood of patients with metastatic breast cancer: a validation study of the CellSearch system. Clin. Cancer Res. 13, 920–928.

Rosenberg, H., Pollock, N., Schiemann, A., Bulger, T., Stowell, K., 2015. Malignant hyperthermia: a review. Orphanet J. Rare Dis. 10, 93.

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., Arnheim, N., 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230, 1350–1354.

Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., Mcgrath, L.M., Kosmicki, J.A., Rehnstrom, K., Mallick, S., Kirby, A., Wall, D.P., Macarthur, D.G., Gabriel, S.B., Depristo, M., Purcell, S.M., Palotie, A., Boerwinkle, E., Buxbaum, J.D., Cook Jr., E.H., Gibbs, R.A., Schellenberg, G.D., Sutcliffe, J.S., Devlin, B., Roeder, K., Neale, B.M., Daly, M.J., 2014. A framework for the interpretation of *de novo* mutation in human disease. Nat. Genet. 46, 944–950.

Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., Dilullo, N.M., Parikshak, N.N., Stein, J.L., Walker, M.F., Ober, G.T., Teran, N.A., Song, Y., El-Fishawy, P., Murtha, R.C., Choi, M., Overton, J.D., Bjornson, R.D., Carriero, N.J., Meyer, K.A., Bilguvar, K., Mane, S.M., Sestan, N., Lifton, R.P., Gunel, M., Roeder, K., Geschwind, D.H., Devlin, B., State, M.W., 2012. *De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 485, 237–241.

Sausen, M., Parpart, S., Diaz Jr., L.A., 2014. Circulating tumor DNA moves further into the spotlight. Genome Med. 6, 35.

Schaaf, C.P., Gonzalez-Garay, M.L., Xia, F., Potocki, L., Gripp, K.W., Zhang, B., Peters, B.A., Mcelwain, M.A., Drmanac, R., Beaudet, A.L., Caskey, C.T., Yang, Y., 2013. Truncating mutations of *MAGEL2* cause Prader-Willi phenotypes and autism. Nat. Genet. 45, 1405–1408.

Scott Jr., R.T., Upham, K.M., Forman, E.J., Zhao, T., Treff, N.R., 2013. Cleavage-stage biopsy significantly impairs human embryonic implantation potential while blastocyst biopsy does not: a randomized and paired clinical trial. Fertil. Steril.

Shah, S., Schrader, K.A., Waanders, E., Timms, A.E., Vijai, J., Miething, C., Wechsler, J., Yang, J., Hayes, J., Klein, R.J., Zhang, J., Wei, L., Wu, G., Rusch, M., Nagahawatte, P., Ma, J., Chen, S.C., Song, G., Cheng, J., Meyers, P., Bhojwani, D., Jhanwar, S., Maslak, P., Fleisher, M., Littman, J., Offit, L., Rau-Murthy, R., Fleischut, M.H., Corines, M., Murali, R., Gao, X., Manschreck, C., Kitzing, T., Murty, V.V., Raimondi, S.C., Kuiper, R.P., Simons, A., Schiffman, J.D., Onel, K., Plon, S.E., Wheeler, D.A., Ritter, D., Ziegler, D.S., Tucker, K., Sutton, R., Chenevix-Trench, G., Li, J., Huntsman, D.G., Hansford, S., Senz, J., Walsh, T., Lee, M., Hahn, C.N., Roberts, K.G., King, M.C., Lo, S.M., Levine, R.L., Viale, A., Socci, N.D., Nathanson, K.L., Scott, H.S., Daly, M., Lipkin, S.M., Lowe, S.W., Downing, J.R., Altshuler, D., Sandlund, J.T., Horwitz, M.S., Mullighan, C.G., Offit, K., 2013. A recurrent germline *PAX5* mutation confers susceptibility to pre-B cell acute lymphoblastic leukemia. Nat. Genet. 45, 1226–1231.

Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B., Hood, L.E., 1986. Fluorescence detection in automated DNA sequence analysis. Nature 321, 674–679.

Steiner, B., Masood, R., Rufibach, K., Niedrist, D., Kundert, O., Riegel, M., Schinzel, A., 2015. An unexpected finding: younger fathers have a higher risk for offspring with chromosomal aneuploidies. Eur. J. Hum. Genet. 23, 466–472.

Veltman, J.A., Brunner, H.G., 2012. *De novo* mutations in human genetic disease. Nat. Rev. Genet. 13, 565–575.

Vissers, L.E., Gilissen, C., Veltman, J.A., 2016. Genetic studies in intellectual disability and related disorders. Nat. Rev. Genet. 17, 9–18.

Wang, J., Wang, W., Li, R., Li, Y., Tian, G., Goodman, L., Fan, W., Zhang, J., Li, J., Guo, Y., Feng, B., Li, H., Lu, Y., Fang, X., Liang, H., Du, Z., Li, D., Zhao, Y., Hu, Y., Yang, Z., Zheng, H., Hellmann, I., Inouye, M., Pool, J., Yi, X., Zhao, J., Duan, J., Zhou, Y., Qin, J., Ma, L., Li, G., Zhang, G., Yang, B., Yu, C., Liang, F., Li, W., Li, S., Ni, P., Ruan, J., Li, Q., Zhu, H., Liu, D., Lu, Z., Li, N., Guo, G., Ye, J., Fang, L., Hao, Q., Chen, Q., Liang, Y., Su, Y., San, A., Ping, C., Yang, S., Chen, F., Li, L., Zhou, K., Ren, Y., Yang, L., Gao, Y., Yang, G., Li, Z., Feng, X., Kristiansen, K., Wong, G.K., Nielsen, R., Durbin, R., Bolund, L., Zhang, X., Yang, H., 2008. The diploid genome sequence of an Asian individual. Nature 456, 60–65.

Wendel, M., Bazhenova, L., Boshuizen, R., Kolatkar, A., Honnatti, M., Cho, E.H., Marrinucci, D., Sandhu, A., Perricone, A., Thistlethwaite, P., Bethel, K., Nieva, J., Heuvel, M., Kuhn, P., 2012. Fluid biopsy for circulating tumor cell identification in patients with early-and late-stage non-small cell lung cancer: a glimpse into lung cancer biology. Phys. Biol. 9, 016005.

Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., Mcguire, A., He, W., Chen, Y.J., Makhijani, V., Roth, G.T., Gomes, X., Tartaro, K., Niazi, F., Turcotte, C.L., Irzyk, G.P., Lupski, J.R., Chinault, C., Song, X.Z., Liu, Y., Yuan, Y., Nazareth, L., Qin, X., Muzny, D.M., Margulies, M., Weinstock, G.M., Gibbs, R.A., Rothberg, J.M., 2008. The complete genome of an individual by massively parallel DNA sequencing. Nature 452, 872–876.

Wong, W.S., Solomon, B.D., Bodian, D.L., Kothiyal, P., Eley, G., Huddleston, K.C., Baker, R., Thach, D.C., Iyer, R.K., Vockley, J.G., Niederhuber, J.E., 2016. New observations on maternal age effect on germline *de novo* mutations. Nat. Commun. 7, 10486.

Worthey, E.A., Mayer, A.N., Syverson, G.D., Helbling, D., Bonacci, B.B., Decker, B., Serpe, J.M., Dasu, T., Tschannen, M.R., Veith, R.L., Basehore, M.J., Broeckel, U., Tomita-Mitchell, A., Arca, M.J., Casper, J.T., Margolis, D.A., Bick, D.P., Hessner, M.J., Routes, J.M., Verbsky, J.W., Jacob, H.J., Dimmock, D.P., 2011. Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. Genet. Med. 13, 255–262.

Yang, Y., Muzny, D.M., Reid, J.G., Bainbridge, M.N., Willis, A., Ward, P.A., Braxton, A., Beuten, J., Xia, F., Niu, Z., Hardison, M., Person, R., Bekheirnia, M.R., Leduc, M.S., Kirby, A., Pham, P., Scull, J., Wang, M., Ding, Y., Plon, S.E., Lupski, J.R., Beaudet, A.L., Gibbs, R.A., Eng, C.M., 2013. Clinical whole-exome sequencing for the diagnosis of Mendelian disorders. N. Engl. J. Med. 369, 1502–1511.

Zhang, J., Walsh, M.F., Wu, G., Edmonson, M.N., Gruber, T.A., Easton, J., Hedges, D., Ma, X., Zhou, X., Yergeau, D.A., Wilkinson, M.R., Vadodaria, B., Chen, X., Mcgee, R.B., Hines-Dowell, S., Nuccio, R., Quinn, E., Shurtleff, S.A., Rusch, M., Patel, A., Becksfort, J.B., Wang, S., Weaver, M.S., Ding, L., Mardis, E.R., Wilson, R.K., Gajjar, A., Ellison, D.W., Pappo, A.S., Pui, C.H., Nichols, K.E., Downing, J.R., 2015. Germline mutations in predisposition genes in Pediatric Cancer. N. Engl. J. Med. 373, 2336−2346.

Zheng, G.X., Lau, B.T., Schnall-Levin, M., Jarosz, M., Bell, J.M., Hindson, C.M., Kyriazopoulou-Panagiotopoulou, S., Masquelier, D.A., Merrill, L., Terry, J.M., Mudivarti, P.A., Wyatt, P.W., Bharadwaj, R., Makarewicz, A.J., Li, Y., Belgrader, P., Price, A.D., Lowe, A.J., Marks, P., Vurens, G.M., Hardenbol, P., Montesclaros, L., Luo, M., Greenfield, L., Wong, A., Birch, D.E., Short, S.W., Bjornson, K.P., Patel, P., Hopmans, E.S., Wood, C., Kaur, S., Lockwood, G.K., Stafford, D., Delaney, J.P., Wu, I., Ordonez, H.S., Grimes, S.M., Greer, S., Lee, J.Y., Belhocine, K., Giorda, K.M., Heaton, W.H., Mcdermott, G.P., Bent, Z.W., Meschi, F., Kondov, N.O., Wilson, R., Bernate, J.A., Gauby, S., Kindwall, A., Bermejo, C., Fehr, A.N., Chan, A., Saxonov, S., Ness, K.D., Hindson, B.J., Ji, H.P., 2016. Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. Nat. Biotechnol.

Chapter 10

Application of Padlock and Selector Probes in Molecular Medicine

I. Hernández-Neuta[1], A. Mezger[1], L. Moens[2], C. Larsson[2], J. Stenberg[2], J. Göransson[2], I. Grundberg[2], M. Isaksson[2], T. Conze[2], S. Henriksson[2] and M. Nilsson[1,2]

[1]*Stockholm University, Solna, Sweden;* [2]*Uppsala University, Uppsala, Sweden*

10.1 INTRODUCTION

One of the rate-limiting factors in the progression of molecular medicine is the techniques available for molecular analysis. Improvements in technology are often followed by scientific breakthroughs, exemplified by techniques such as molecular cloning, DNA sequencing, and polymerase chain reaction (PCR). A common objective in molecular medicine is to analyze the quantity and/or sequence of a large number of candidate genes or loci of genetic variation. The number of genes under study is often significantly larger than one and yet significantly fewer than all. This intermediary range in the analytical spectrum is technically challenging. Techniques that perform well when analyzing individual targets do not automatically work well when targeting multiple genes. Conversely, techniques that enable the global analysis of virtually all genes are not cost-effective and accurate enough to be implemented directly for targeted analyses, particularly when aiming for diagnostics. In this chapter some approaches for targeted multiplex analysis for genotyping and massively parallel sequencing are described.

Another important area of investigation is to study where in a complex tissue a particular gene or gene variant is expressed. Here, existing in situ hybridization techniques do not readily resolve single-nucleotide variants, or even splice variants of genes. New techniques are needed to improve the resolution of such single-cell analyses in the intact context of tissue samples.

In this chapter, we will describe assays based on padlock and selector probes, both employing DNA-ligase assisted DNA circularization reactions. We will describe some unique properties of these circularization approaches and the advantages they bring. The assays have been used for large-scale genotyping and sequencing of targeted parts of the genome, DNA copy number analysis, infectious diagnostics, single-molecule detection, and in situ analysis with high multiplex ability and single-nucleotide resolution. The different applications and techniques will be described in the following sections.

10.2 PADLOCK AND SELECTOR PROBES

Both padlock and selector probe assays are based on highly specific DNA ligase-catalyzed DNA circularization reactions (Figs. 10.1 and 10.2). Even if the overall architecture of the assays is similar, the outcome of the circularization reaction is different because ligation of padlock probes is templated by the target sequence, whereas it is the other way around in the selector probe assay, as described subsequently. Circularization can be performed by ligation only, after a gap-fill reaction or after an invasive cleavage reaction (Figs. 10.1 and 10.2).

10.2.1 Padlock Probes

The padlock probe concept was invented by Ulf Landegren in the early 1990s as a development of his oligonucleotide ligation assay (Landegren et al., 1988) and was published in 1994 (Nilsson et al., 1994). A padlock probe consists of three segments: two target complementary end-sequences and a nontarget complementary element that links the end sequences together. The target-complementary segments are typically in the range of 15−22 nucleotides in length and the linking segment is typically longer than the combined length of the target-complementary segments. The

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00010-9
Copyright © 2017 Elsevier Ltd. All rights reserved.

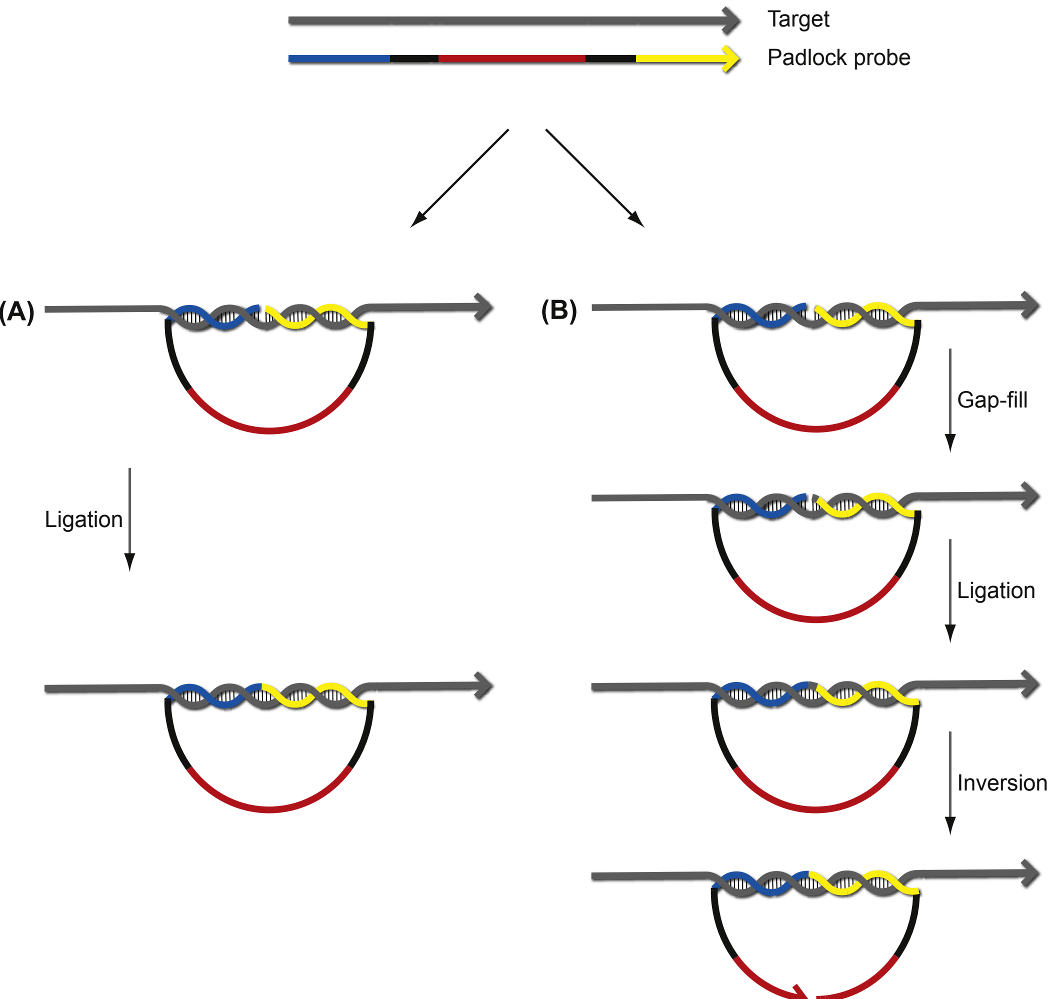


FIGURE 10.1 The architecture of the padlock and molecular inversion probe assays. Both assays employ probe molecules consisting of 5′-target complementary sequences (blue), a linking segment (black) including sequences for amplification and identification (red), and a 3′-target-complementary sequence (yellow). The ends bind head-to-tail when the probe hybridizes to a denatured single-stranded target molecule (gray), forming a structure that can be circularized through the action of one or several enzymatic reactions. (A) In the original padlock probe assay, the ends are joined by a DNA ligase. (B) In the molecular inversion probe assay, a single-nucleotide gap is filled with a DNA polymerase to create a suitable structure for the DNA ligase-assisted probe circularization. After an exonuclease clean-up of the sample, degrading linear target and probe molecules, thereby enriching for the circularized molecules, the probe molecules are opened by degradation of uracil residues in the linking segment. The linearization serves to release the probe molecules from the encircled target molecules, improving the amplification efficiency, and may also reduce probe-dependent amplification artifacts.

linking segment can also be composed of a polymer other than DNA, e.g., polyethylene glycol (Nilsson et al., 1994). The padlock probe design is such that the ends of the probe hybridize head to tail on the single-stranded target molecule, forming a nicked DNA duplex that can be sealed by a DNA ligase (Fig. 10.1A). The linking segment typically harbors sequences for amplification (primer motifs) and identification of the amplification product (tag sequences). Amplification of padlock probes can be done using PCR, rolling circle amplification (RCA), and exponential variants of RCA [hyperbranched RCA (hRCA) and circle-to-circle amplification (C2CA)] or combinations of them, which will be described in the following sections.

10.2.2 Gap-Fill and Molecular Inversion Probes

In the gap-fill variation of the padlock probe, the ends are designed to hybridize a distance apart on the target. The gap between the ends is then filled in by a polymerase before circularization by ligation (Akhras et al., 2007; Hardenbol et al., 2003; Porreca et al., 2007) (Fig. 10.2C). The gap-fill padlock probes are described in more detail in Section 10.6, where techniques for multiplexed targeted resequencing are discussed. In the molecular inversion probe (MIP) variant of gap-fill padlock probes, the gap consists of a single nucleotide at a single-nucleotide

FIGURE 10.2 The architecture of the selector and gap-fill padlock probe assays. (A) Similar to padlock probes, selector probes consist of a 5′-target-complementary sequence (blue), a linking segment containing sequences for universal amplification reagents (black), and a 3′-target complementary sequence (yellow). The linking segment is made double-stranded by hybridization to a complementary vector oligonucleotide (black). The partially double-stranded selector probe acts as a template for a circularization of a specific genomic DNA fragment through two alternative routes. In the simplest format the selector probe hybridizes to the ends of a denatured single-stranded restriction fragment containing the target sequence, forming a structure that can be circularized by two ligation reactions. (B) Alternatively, the 5′-end of a restriction fragment can be cleaved in a structure-specific endonucleolytic "invasive" reaction to minimize the nontarget sequence content of a restriction fragment. After that reaction, a structure is formed that is identical to the one in (A) that is circularized by two ligation reactions. (C) The gap-fill padlock probe ligation reaction acts through a completely different mechanism in which the target strand is first replicated to fill the gap between the two target-complementary probe sequences of the gap-fill probe before a structure is formed that can be circularized through the action of a DNA ligase. This approach does not require restriction digestion, which simplifies the assay design.

polymorphism (SNP) position. The SNP is queried in four separate tubes, each containing one of the deoxynucleotides (dNTPs), a DNA polymerase, and a DNA ligase (Hardenbol et al., 2003). Circularization is used in conjunction with enzymatic cleavage of the linking segment, resulting in the reacted probe being converted into a linear molecule with the order of the segments inverted (Fig. 10.1B). At the same time, unreacted probes are cleaved in two parts, which reduces the risk of background signal in a subsequent amplification reaction. The MIP has been developed to construct highly multiplexed targeted genotyping assays (see Section 10.3).

10.2.3 Selector Probes

A selector probe resembles a padlock probe in that it consists of three segments and the end segments are complementary to a genomic target sequence. However, these ends are

designed to hybridize to each of the ends of a single-stranded restriction fragment. Also, a general vector oligonucleotide is hybridized to the internal segment of the selector probe, allowing two ligation events to close the construct into a circle (Fig. 10.2A). The circle is then composed of the targeted genomic sequence, and the vector sequence. As an alternative, the 5′-part of the restriction fragment can be trimmed before ligation. This is performed by hybridizing the 5′-end of the selector probe to an internal part of the target fragment. A structure is thus formed that can be cut by a structure-specific endonuclease, e.g., Taq polymerase, as described by Lyamichev et al. (1993), in which the protruding 5′-part of the fragment is released (Fig. 10.2B).

10.3 APPLICATION OF PADLOCK AND MOLECULAR INVERSION PROBES FOR GENOTYPING

Padlock probes are well-suited for targeted multiplex analysis of genetic variation. The probes require dual recognition of the target sequence to become circularized, providing sufficient specificity to probe unique sequences in the genome (Antson et al., 2000; Lizardi et al., 1998). Robust discrimination of single-nucleotide differences is achieved as a result of the stringent substrate requirement of the DNA ligase used in the circularization reaction (Landegren et al., 1988). Finally, the unimolecular circularization approach renders padlock probe-based multiplexed assays less prone to problems with cross-reactivity than bimolecular assays such as multiplexed PCR. Therefore a large number of circularization probes can be combined in multiplexed assay without the problem of interference (Banér et al., 2003; Hardenbol et al., 2003). The probes are equipped with sequence motifs in the linking segment that are used to sort the amplification product from a multiplex padlock probe assay on a microarray comprising tag-sequences that are designed to be maximally divergent in sequence. The specificity of such tag−array hybridizations is typically good enough for genotyping, but when there are differences in the concentration of different targets, such as in transcription profiling, there are limitations in the dynamic range of this analysis platform. This limitation has been overcome by introducing a ligation-based tag-recognition coupled to localized RCA signal amplification (Ericsson et al., 2008).

The padlock probe in its original design has been used in moderately multiplexed assays for infectious disease diagnostics (see Section 10.5.1), in situ genotyping and detection assays (see Section 10.8), and genotyping (Banér et al., 2003; van Eijk et al., 2004), and to determine which T-cell receptor is expressed in clonally expanded T-cell populations by probing for the expression of the entire repertoire of Vb genes (Banér et al., 2005).

10.3.1 Genotyping Using Padlock Probes and Hyperbranched Rolling Circle Amplification

The surprising finding that small DNA circles can also template rolling circle DNA synthesis was shown independently by Fire and Xu (Fire and Xu, 1995) and Kool et al. (Liu et al., 1996). Here, the DNA polymerase replicates a circular DNA template repeatedly, building a long single-stranded tandem repeated product. Padlock probes were coupled early to this RCA mechanism (Banér et al., 1998; Lizardi et al., 1998; Zhang et al., 1998), which has been critical to many of the applications discussed later in this chapter. Φ29 DNA polymerase has been shown to be particularly well-suited for RCA, perhaps because of the combination of strand displacement activity, high processivity, and both single-stranded and double-stranded 3′-exonucleolytic (proofreading) activity (Banér et al., 1998; Lizardi et al., 1998; Salas et al., 2004). One of the first applications of padlock probe RCA used an exponential version of the technique called hRCA (Banér et al., 1998; Lizardi et al., 1998; Zhang et al., 1998). Here, the RCA is performed with a forward and a reverse primer, just like in PCR, but in this case every primer binding event gives rise to multiple copies of the tandem-repeated RCA product and the process is isothermal because the product strands are separated through the strand displacement activity of the DNA polymerase. The same mechanism, coupled with random hexamer primer libraries, is used in the multiple strand displacement amplification employed to amplify the content of DNA samples uniformly (Dean et al., 2002). The amplification is exponential, allowing for rapid detection of SNPs. Genotyping assays have been developed based on different incarnations of this technique, using PCR products or genomic DNA as template, and calling the alleles in separate tubes using intercalating dyes or in one tube using primers that emit fluorescence when consumed in the reaction (Alsmadi et al., 2003; Faruqi et al., 2001; Pickering et al., 2002; Thomas et al., 1999).

10.3.2 Massively Multiplex Targeted Genotyping Using Molecular Inversion Probes

One significant application of circularization probes the massively multiplexed targeted genotyping that has been achieved with the MIPs. The method was initially developed at Stanford by Hardenbol et al. (2003) and then was translated to 10,000-plex assays at the company ParAllele (Hardenbol et al., 2005), which was later acquired by Affymetrix (http://www.affymetrix.com). The technique has been used in a variety of applications, as briefly

described subsequently. A more detailed review of the MIP assay was compiled by Absalan and Ronaghi (2007). The assay is available as kits of probes and tag-microarrays to obtain genotyping results from 1000 to 25,000 targeted SNPs. The technique has been used in the human and bovine HapMap projects (Consortium, 2005; Frazer et al., 2007; Gibbs et al., 2003; Khatkar et al., 2007) to establish a map of the most common haplotype blocks in these populations and to investigate the extent of linkage disequilibrium in populations of wild mice (Laurie et al., 2007). The MIP assay has also been used in association studies to map genes' increasing susceptibility to type 1 diabetes (Smyth et al., 2006), prostate cancer (Zheng et al., 2006), and colorectal cancer (Zanke et al., 2007). The assay is also suitable for analysis of degraded DNA, such as is found in formalin-fixed paraffin-embedded (FFPE) tissue samples, because of the short footprint of the probes and the fact that the hybridization and ligation reactions are probably more tolerant to DNA damage than are techniques that require replication of the target strand, such as PCR (Wang et al., 2005). These authors also demonstrated the technique's use in detecting copy number variations (CNVs) in the sample. The assay is now commercialized by Affymetrix as its OncoScan FFPE Express Service, which offers high-quality copy number and genotype data with whole-genome coverage and high resolution in cancer genes for use with challenging FFPE samples (Wang et al., 2012).

10.4 BIOSENSOR APPROACHES BASED ON ROLLING CIRCLE–AMPLIFIED PADLOCK PROBES

The combination of the strictly target-dependent padlock probe circularization reaction and the strictly circle-dependent RCA reaction offers opportunities for novel biosensor approaches enabling single-molecule detection in multiplexed assays. Some of these approaches will be described in this section.

10.4.1 Homogenous Amplified Single-Molecule Detection

Most methods used for molecular analyses measure an average of a molecular population, for instance, based on the fluorescence intensity of a reporter molecule. To reach ultimate precision and quantification, however, a digital readout is required, enabling the counting of individual molecules. Current techniques for direct single-molecule detection (SMD), especially ones with fluorescence readout of the results, demand highly advanced equipment and are often slow and not robust.

An alternative approach is based on amplified SMD, in which individual molecules are clonally amplified and detected. Detection of single templates with PCR was first demonstrated by Li et al. (1988), in which single human sperm cells were added to individual reactions to allow the detection of recombination events in the male germline. A decade later, Vogelstein and Kinzler (1999) used single-molecule PCRs to achieve digital quantification of somatic mutations. Instead of recording the degree of amplification, the reactions were treated as binary calls ("yes" or "no" answer), and the technique was therefore called "digital PCR." Analysis of single DNA molecules was achieved by diluting the sample in microtiter plates to the extent that on average, only every second well contained a template. Digital PCR has been further developed by replacing the well compartments with oil–water micelle microreactors containing microbeads enabling higher throughput. With this strategy, the locally amplified products are fluorescently labeled according to their target sequence and analyzed with flow cytometry (Dressman et al., 2003). However, emulsion PCR is mainly used as an amplification method to create single-molecule clones as templates for several next-generation sequencing strategies.

Another way to transform a population of DNA molecules to amplified single molecules, and thus preserve its digital nature, is to use RCA. Interestingly, the long DNA threads produced in the RCA collapse into a random coil with a diameter of about 700 nm, creating inherently localized DNA clusters (Melin et al., 2007). A local enrichment of fluorescence up to a factor of 1000 can be achieved in a homogeneous format by hybridizing fluorescence-labeled oligonucleotide probes to the repeats in the rolling-circle product (RCP), resulting in bright fluorescent objects that are easily visualized using standard fluorescence microscopy (Blab et al., 2004). The high signal-to-noise ratio allows robust and rapid detection and counting of amplified molecules by pumping the RCP solution through a microfluidic channel mounted on a confocal microscope operating in a scanning mode across the width of the flow channel (Jarvius et al., 2006) (Fig. 10.3).

A padlock probe is usually equipped with a tag, correlating the reacted probe to its target sequence. Hence, different RCP populations can be labeled with different types of fluorescence and spectrally distinguished. In fact, amplification by RCA is well-suited for multiplexing because the RCPs rarely interact with each other, and even fully complementary RCPs aggregate to a negligible degree (Melin et al., 2007). The approach of counting individual RCPs homogeneously has been used to detect and quantify different pathogens selectively, as described in Section 10.5.1.

RCPs can be immobilized on different solid supports or generated by a localized amplification reaction to obtain surface-confined digital quantification objects. This approach has been demonstrated on flat surfaces generating highly multiplex random single-molecule arrays (Göransson et al., 2009)

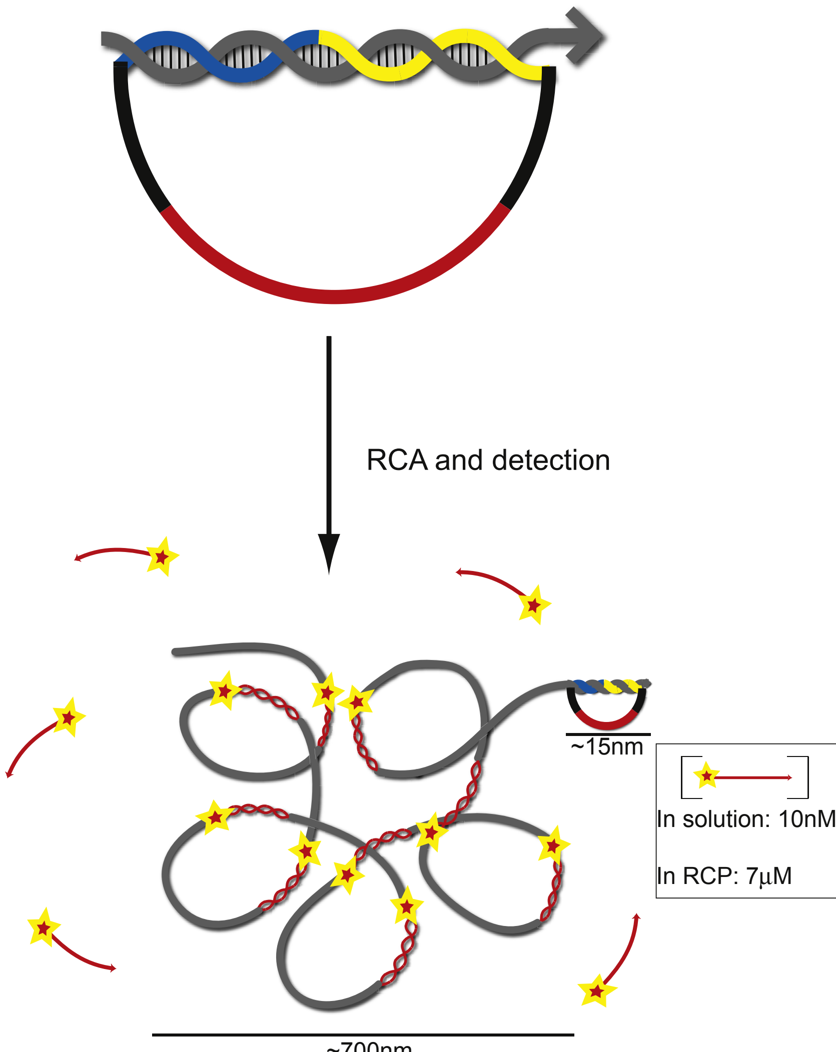


FIGURE 10.3 Padlock probe and rolling circle amplification (RCA)-based single-molecule detection. A padlock probe is circularized in a strictly target-dependent ligation reaction. Then a long tandem-repeated DNA product is formed in a strictly circle-dependent RCA reaction, thus forming a rolling-circle product (RCP) in a strictly target-dependent manner. The RCA product is labeled with fluorescence-tagged short oligonucleotide probes (red with yellow star) that hybridize to a tag sequence in the product (red). The long product strand collapses in to a submicron-sized object owing to random coiling. The local concentration of fluorescence tags in a product composed of 1000 copies of the circle is about 7 μM, compared with 10 nM in the surrounding solution, which is sufficient to saturate the RCPs with fluorescence tags. This difference in concentration makes it possible to detect single RCPs in solution with no need to wash away excess fluorescence tags.

and on beads (Sato et al., 2013; Schopf et al., 2011). This principle has found different applications such as detecting rare enzymatic events for theranostics (Juul et al., 2011) and the sensitive detection of low-abundance surface-marker proteins (Konry et al., 2011).

In an interesting approach, individual label-free RCPs were counted using a resistive pulse-sensing nanopore (Kühnemund and Nilsson, 2014). Voltage is applied through a nanopore that is placed between two compartments, causing an ionic current that can be affected by particles that pass through the nanopore. The amplitude and the width of the current pulse are measured, providing information about the size and charge of the particle in a fast and simple setup (Roberts et al., 2010; Vogel et al., 2011). RCPs attached to beads were detected and quantified with this approach by distinguishing between the sizes of the particles that had attached RCPs from blank particles. This was the first time that reacted padlock probes were detected with a free label scheme.

10.4.2 Magnetic Biosensor Approaches

Magnetic nanoparticles, commonly referred to as magnetic beads, are applied in many research fields including biotechnology, biomedicine, and drug discovery (Saiyed et al., 2003). Several bioassays have been developed based on the Brownian relaxation biosensor scheme. Here, the susceptibility of magnetic beads to reorient in an AC field is measured. The smaller the bead diameter, the more rapid is its response to a shift in the electric field. This response has a characteristic critical frequency, called the Brownian relaxation frequency. In these biosensor approaches, small magnetic beads are equipped with probes, and upon interaction with a target a slight decrease in relaxation frequency can be observed as an effect of the increased hydrodynamic volume of the magnetic bead. The frequency shift can be recorded with a magnetometer, a superconducting quantum interference device. In a study by Strömberg et al. (2008), the Brownian relaxation principle was used to detect padlock probe-generated RCPs. RCPs bound to magnetic beads cause a dramatic shift in frequency because of their large size and multivalent binding interactions between the beads and the RCPs. This effect can easily be observed as a loss of signal in the free-bead frequency domain and the response can be quantified by measuring the decrease in amplitude, which correlates well with the RCP concentration. This in turn corresponds to the number of reacted padlock probes, and therefore to the target concentration. With the magnetometer used in the study, the limit of detection was in the low picomolar range. Multiplex detection was achieved using different sized particles functionalized with tag-specific oligonucleotides. A biplex assay was demonstrated and the possibility of increasing the number of targets was shown (Strömberg et al., 2009).

This approach took a step closer in its direct applicability when a portable AC susceptometer was proven to be efficient for measuring the frequency shift. A limit of detection of 4 pM was reported and results were obtained in less than 30 min at room temperature (de la Torre et al., 2011; Strömberg et al., 2014). Another example of the use of this principle was developed by Østerberg et al. (2013, 2014), who simplified the way to perform the magnetic read-out by using magnetoresistive planar Hall effect bridge sensors that can measure the hydrodynamic size of magnetic nanobeads without the need for external magnetic fields or bulky equipment. Alternatively, instead of measuring the frequency shift directly, it is possible to detect the response of nanobeads after applying an AC magnetic field by measuring the modulation of the intensity of light transmitted through the magnetic bead dispersion. This approach significantly simplified the technical requirements needed to perform this type of measurement (Bejhed et al., 2015; Donolato et al., 2015).

10.4.3 DNA Nanowires

DNA threads can nucleate metal nanoparticle deposition creating electrically conductive nanowires. Therefore, the presence of a certain DNA sequence can be sensed by producing such structures and measuring the electrical resistance in a given circuit. RCPs, which are long single-strand DNA (ssDNA) threads, can be metallized with gold or silver nanoparticles. Russell et al. (2014) reported a biosensor based on the formation of nanowires seeded by stretched RCPs across two electrodes. These RCPs have been specifically formed on one electrode after target-directed padlock probe ligation. The amplified molecules are then stretched by drying the surface in a directed flow, connecting the two separated electrodes. Metal nanoparticles functionalized with complementary oligonucleotides are then hybridized to the stretched RCPs. These particles will seed the deposition of metal from solution all along the DNA molecule, forming a metal nanowire that bridges two electrodes and thus causes a drop in the resistance of the electrical circuit. With this approach, 10 ng of bacterial DNA was successfully detected.

10.5 APPLICATION OF PADLOCK PROBES FOR INFECTIOUS DISEASE DIAGNOSTICS

A common task in infectious disease diagnostics is to identify an infectious agent from a panel of candidate pathogens. Padlock probe assays are well-suited for these types of analyses owing to their multiplexing ability and specificity. Compared with culture-based diagnostics, these assays accelerate diagnosis from days or weeks to hours (Kong et al., 2008). Padlock probes and similar technologies have been successfully applied for the detection of a wide variety of pathogens in various assay formats. The assays are directed to RNA or DNA target sequences and use real-time fluorescence, microarray, SMD, and colorimetric or luminescence readouts. Some yield sequence information whereas others quantify pathogens. Padlock probes have also been applied for in situ analysis of infectious organisms, as described in Section 10.8.4. In past years, thanks to numerous molecular and technological developments, applications for padlock probes for pathogen detection have been demonstrated in different fields, from agriculture to veterinary medicine to clinical microbiology, as will be described subsequently (Kong et al., 2008; Tong et al., 2007).

10.5.1 Padlock Probe-Based Infectious Disease Diagnostics

The potential of padlock probe assays for infectious disease diagnostics has been shown by targeting different etiologic

agents that range from bacteria to viruses in different applications and formats. Important viral livestock pathogens were detected and genotyped using multiplexed padlock ligation to cDNA, employing tag-microarrays to analyze the RCA and PCR-amplified padlock probes (Banér et al., 2007). The padlock probes were designed against several highly conserved regions in each virus. This redundant testing with independent probes renders the assay tolerant to novel mutations. The same principle of multiplexed padlock probe ligation to cDNA was applied by Gyarmati et al. (2008), who developed a padlock probe assay to genotype all 16 hemagglutinin and all 9 neuraminidase gene subtypes of avian influenza virus. All known isolates of the avian influenza virus were correctly subtyped in single reactions. Influenza A H1N1 subtype, the causative agent of an influenza pandemic in 2009, and mutations causing antiviral resistance have also been detected with padlock probes and hRCA with real-time fluorescence detection in different studies (Steain et al., 2009; Wang et al., 2010).

van Doorn et al. (2007) combined multiplex padlock probe assays with quantitative PCR (qPCR) readout in a novel assay format to enable an increased dynamic range compared with a microarray readout. A pool of padlock probes directed to 11 plant pathogens were reacted with the samples in a single reaction and then reversibly bound to beads. The target sequences were then washed off from the beads, the probes were eluted, and nonreacted probes were removed by exonucleases. The remaining circular padlock probes were loaded onto an OpenArray qPCR platform that allows for 3072 PCRs with up to 64 different primer sets in 33-nL wells (http://www.biotrove.com). Each pathogen-specific padlock probe was equipped with a specific PCR−primer pair motif and was analyzed in separate qPCR reactions.

RNA viruses have also been detected with padlock probes without reverse transcription into cDNA (Millard et al., 2006). Diagnostic applications outside a laboratory environment are enabled by biosensor devices that perform padlock probe ligation, isothermal amplification, and detection reactions with fluorescence readout (McCarthy et al., 2006, 2007).

An important development that contributed to expand the field of application of padlock probe-based assay was the combination of C2CA (Dahl et al., 2004) with magnetic microbeads (Ke et al., 2011). These solid supports with a large surface provided a practical way to concentrate and isolate targets from complex samples and enabled washing of excess padlock probes, enzymes, and other reagents. This process greatly simplified the assay configuration by favoring the kinetics of the reaction, making it faster and more robust (Göransson et al., 2012; Ke et al., 2011). This magnetic bead−based C2CA (Fig. 10.4) has proven to be advantageous because it constitutes a PCR-free format and thus diminishes the technical requirements needed to be performed. Adaptation of this assay for the detection of infectious diseases has led to the development of novel biosensor techniques and opened up the possibility of applying padlock probe-based pathogen detection in a clinical setting.

Ke et al. (2011) used a cocktail of padlock probes and magnetic bead−based C2CA to detect the Crimean-Congo hemorrhagic fever virus (CCHFV). Using a cocktail of padlock probes, different CCHFV strain variants were specifically detected. A sensitivity of 10^3 viral copies/mL was reached by implementing a 96-well plate readout format with colorimetric detection using horseradish peroxidase chemistry, which is widely used in clinical diagnostics in enzyme-linked immunosorbent assays. Similarly, C2CA in conjunction with SMD was applied to detect rotavirus directly in clinical samples. Degenerated bases were introduced at the arms of the padlock probes, creating a cocktail of probes and covering 95% of the published rotavirus A sequences in the National Center for Biotechnology Information database. Remarkably, a total assay time of 3.5 h was achieved, with a potentially higher specificity compared with routinely used diagnostic techniques such as latex agglutination assay or chromatography immunoassay and a better multiplex ability than qPCR-based assays (Mezger et al., 2014).

Various diagnostically relevant bacteria have been successfully detected by applying padlock probes, C2CA, and SMD to target DNA directly. The specificity of padlock probes has been particularly advantageous for detecting drug resistance-associated genes, plasmids, and single-nucleotide variations, which is advantageous for developing diagnostic assays requiring reliability in clinical practice for accurate drug prescription.

As an example, Engström et al. (2013) reported a multiplex C2CA assay using a magnetic readout able specifically to detect the highly prevalent and slow-growing pathogen *Mycobacterium tuberculosis* and to discriminate between wild-type and nine different single-nucleotide variants associated with rifampicin resistance, which has been challenging for other PCR-based assays. Using a similar assay configuration, rapid antibiotic susceptibility testing and bacterial identification in urinary tract infections were described (Mezger et al., 2015a). This work combined bacterial culture with a C2CA-based molecular assay detecting *Escherichia coli*, *Pseudomonas aeruginosa*, and *Proteus mirabilis*. Clinical samples were briefly cultured in the presence or absence of antibiotics; then a simple DNA extraction method, padlock probe specific ligation and C2CA, was performed with a total assay time of 3.5 h. This was tested in 88 clinical samples and a sensitivity and specificity of 100% were obtained.

In addition, some applications have been reported for detecting pathogenic fungus. By combining real-time qPCR measurements with a bead-based suspension

FIGURE 10.4 Magnetic bead (MB)−based circle-to-circle amplification. Nucleic acids are first prepared (left). Then extracted RNA is converted to complementary DNA by reverse transcription using a biotinylated primer that can be captured to magnetic beads coated with streptavidin. In the case of genomic DNA, this is fragmented by a restriction enzyme or other shearing method and then denatured to obtain single-stranded DNA. An excess of padlock probes is added together with a ligase and a capture oligonucleotide functionalized with biotin (in the case of DNA) specifically to capture and detect the region of interest sequence (1). MB functionalized with streptavidin are used to extract/capture circularized probes (2). MB are collected with a magnet and washed to perform a subsequent RCA reaction (3). MB are collected once again and the solution is exchanged for a restriction mixture that contains replication oligonucleotides that have a restriction sequence that upon hybridization will be recognized by a restriction endonuclease causing the digestion of the rolling-circle product (RCP) concatamer, generating monomers that can be recovered in the supernatant (4). These new monomers can form new circles after ligation, and then a second round of RCA is performed (5). These RCPs can be fluorescently labeled for digital single-molecule counting or restriction digested again for use in conventional nucleic acid arrays.

microarray (Luminex™), it was possible to develop a padlock probe assay that detects 10 different human pathogenic fungi with high specificity and a wide dynamic range (Eriksson et al., 2009). This combination has been particularly advantageous because bead arrays diminishes the time needed for hybridization and the fluorescent-encoded beads allow the detection of up to 100 targets simultaneously in a high-throughput format. Mezger et al. (2015b) combined C2CA with this readout format for pathogen detection, once again showing the potential for developing panels that can target multiple pathogens in a single assay.

Detection of pathogens in environmental samples has been achieved by including a desthiobiotin moiety and an internal endonuclease IV cleavage site in the padlock probe linker sequence. These two functionalities served to select the ligated padlock probes and decreased the background of the assay, achieving a sensitivity of 1 pg of DNA and a dynamic range of 10^4 (van Doorn et al., 2009). Furthermore, padlock probes and RCA have been reported to be suitable for screening large numbers of samples. Because of their simplicity and versatility, padlock probes and RCA assays have the potential to be implemented in routine laboratories much more easily than well-known methods

such as amplified fragment-length polymorphism and multilocus sequence typing (Ahmed et al., 2014; Feng et al., 2013; Najafzadeh et al., 2011).

10.5.2 Gap-Fill and Molecular Inversion Probe–Based Infectious Disease Diagnostics

Gap-fill probes have also been applied in pathogen detection and subtyping assays. Novais et al. (2008) developed an assay for the sensitive detection of *M. tuberculosis*. They amplified circularized MIPs by PCR and sequenced the product using pyrosequencing. They further improved sensitivity to a limit of detection of 200 genomes by adding all dNTPs to the bioluminescence reaction performed in the pyrosequencing instrument instead of adding the four nucleotides sequentially, which is required to obtain sequence information. Akhras et al. (2007) developed a gap-fill padlock probe assay, known as connector inversion probe assay, similar to the one described in more detail in Section 10.7.2. DNA circles are produced in an extensive gap-fill polymerization and ligation reaction, allowing the amplification and sequencing of DNA between the two target-complementary segments of a padlock probe. This technology was applied for antibiotic resistance screening in *Neisseria gonorrhoeae* and genotyping of human papillomavirus samples. The same group subtyped human papillomavirus sequences with 24 PCR-amplified MIPs sorted on a tag-microarray (Akhras et al., 2007).

10.6 TARGETED MULTIPLEX COPY NUMBER VARIATION ANALYSIS USING SELECTOR PROBES

Structural variations including deletions, insertions, and inversions have received a lot of attention in the field of genomics. The impact on mammalian phenotypic variation owing to CNV appears to be more extensive than expected (Eichler et al., 2007; Kidd et al., 2008). New methods are required to detect and analyze this variation, which is quantitative in nature, compared with SNPs that basically require only qualitative binary information to score.

10.6.1 Multiple Ligation-Dependent Genome Amplification Assay

Selector probes have been applied to interrogate CNVs using the multiplex ligation-dependent genome amplification (MLGA) assay (Isaksson et al., 2007). MLGA uses selector probes to equip specific genomic fragments with a common primer pair. The selected target fragments can be designed in silico to differ in length and facilitate a capillary gel electrophoresis readout. The relative peak area of the amplification product represents target abundance; by normalizing peaks to reference loci, CNV can be analyzed between a sample and a reference (Fig. 10.5). The MLGA assay can be used to interrogate up to 30–40 fragments, depending on the electrophoretic resolution. They can be used to resolve boundaries of CNVs by spreading the targets for the assay over a CNV candidate region. In a study by Salmon Hillbertz et al. (2007), selector probes were designed to cover a 750-kb region on canine chromosome 18 associated with the Ridge phenotype in Rhodesian Ridgeback dogs. Its name describes its most unique characteristic, a clearly defined symmetrical ridge running the length of its back, formed by hair growing in a direction opposite the rest of its coat. In this duplication boundary mapping approach, a first set of MLGA probes was designed with 100-kb intervals covering a region of 2 Mb. The results confirmed the CNV, and to further map the start and end positions of the region, some additional MLGA probes were designed between duplicated and nonduplicated regions. The borders of the potential tandem repeat were then defined precisely enough to amplify the breakpoint by PCR. The 133.4-kb duplication involved five genes, of which three fibroblast growth factor genes are strong candidates for causing the Ridge phenotype (Salmon Hillbertz et al., 2007).

10.7 HIGH-THROUGHPUT TARGETED SEQUENCING USING SELECTORS AND GAP-FILL PADLOCK PROBES

The past few years have seen accelerating development within methodology and instrumentation for very–high-throughput sequencing (Mardis, 2008). This development will enable new types of genetic analyses such as whole-genome sequencing, deep sequencing of tumor samples, global expression analyses, sequencing of material from chromatin immune-precipitation experiments, and more.

One application to which this new technology is well-suited is the resequencing of many selected parts of a genome, such as all exons from a large set of genes. This requires the targeted parts of the genome to be somehow enriched in the sample. Traditionally this enrichment has been performed using PCR, but this is poorly suited for multiplexing, meaning that a large number of separate reactions would be required. The ideal method for preparing samples as templates for resequencing would be one that can be performed in a single reaction, that enriches all regions of interest (and nothing else), and that yields uniform enrichment for all targeted regions. Attempts to achieve this have been demonstrated using capture and release of selected parts of the genome on custom microarrays, followed by an adaptor-linker PCR (Albert et al., 2007; Okou et al., 2007). In this section, we will describe probe-based

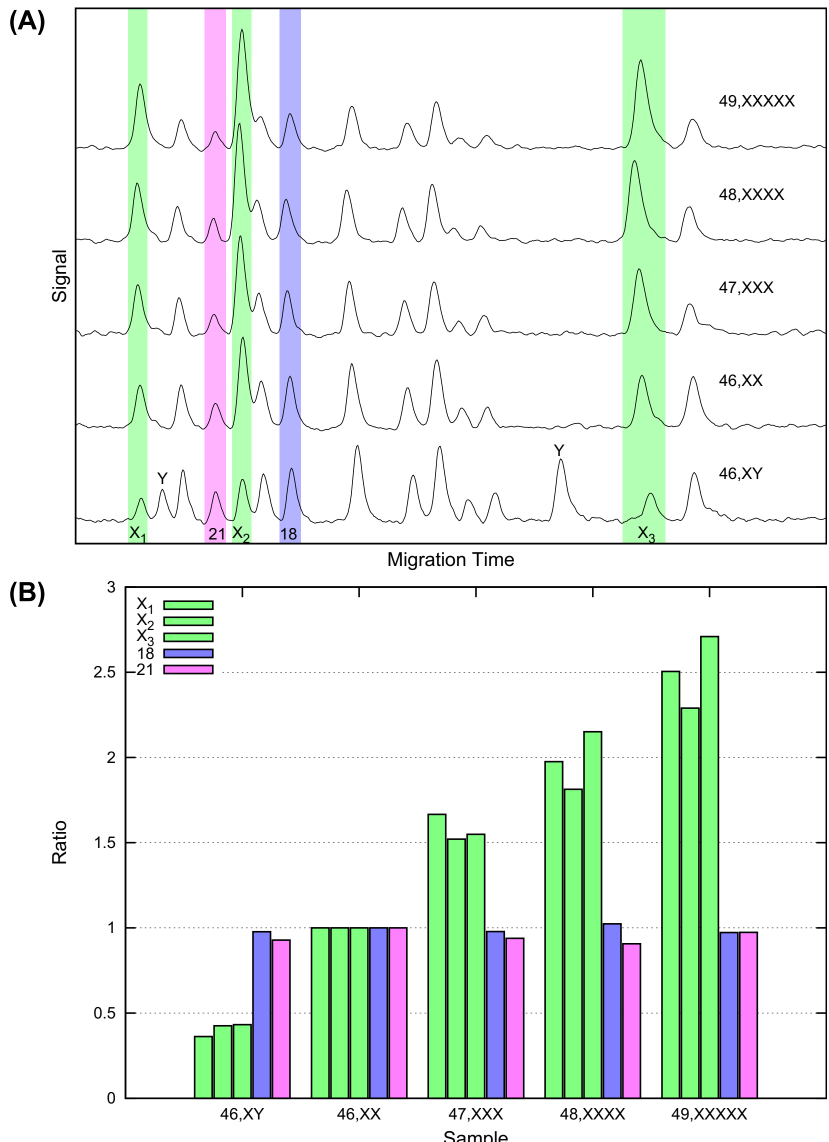


FIGURE 10.5 Multiplex ligation-dependent genome amplification (MLGA) assay for copy number analysis. In the MLGA assay, selector probes are designed to generate amplification products of different size from loci in the genome to be probed for copy number variation. Products from such multiplex amplifications are separated through electrophoresis and the relative amount of each amplification product is measured. Because the relative representation of all amplicons is reproducible, it is possible to measure relative copy number deviations between samples. (A) Elution diagrams from an Agilent Bioanalyzer instrument for five different samples from human cell lines harboring different copy numbers of the X-chromosome. The products from loci located on the X-chromosome are highlighted green and the products from two probes targeting chromosomes 18 and 21 are highlighted blue and red, respectively. Note the presence of two products from the Y-chromosome in the male (XY) sample only and the steady relative increase in the X-chromosomal peaks in the series of one (bottom trace) to five (top trace) X chromosomes. (B) Analysis of the data shown in (A) to determine relative copy number deviations. The areas of the highlighted peaks in (A) were normalized to the peak areas of the normal female sample (46, XX). Two copies of a locus will thus yield a ratio of 1 and three copies 1.5, etc.

methods combining hybridization with enzymatic activity to select the desired parts of the genome.

10.7.1 Targeted Sequencing Using Selector Probes

Selector probes can be employed to amplify a large set of arbitrary restriction fragments in multiplex. These fragments can be selected so that certain desired sequences, e.g., all exons from a set of genes, are amplified. By using endonucleolytic cleavage, as described in Section 10.2.3 (Fig. 10.2B), restriction fragments can be trimmed to decrease the amount of undesired sequence that is also amplified. Dahl et al. (2007) showed for the first time that this method could be combined with massively parallel pyrosequencing (Margulies et al., 2005), to amplify and sequence all coding sequence and flanking regions from 10 genes implicated in colorectal cancer. Approximately 90% of all sequenced reads were from the targeted regions, and 93% of the desired sequence was sequenced at least once. The quality of sequence was sufficient to identify most previously known mutations in the analyzed cell lines.

The selector method (now commercially available as the "HaloPlex target enrichment system," Agilent Technologies) has been used for a wide variety of applications, including the somatic mutation detection of cancer genes (Johansson et al., 2011; Kim et al., 2014; Moens et al., 2015), the diagnosis of primary immunodeficiency diseases (Moens et al., 2014; Stoddard et al., 2014), and population-based association testing in Alzheimer disease (Cuyvers et al., 2015).

Moens et al. (2015) optimized and evaluated the technology for targeted resequencing in clinical FFPE samples with regard to challenges related to DNA-fragmentation, small DNA yields, and formalin-induced base damage. By reducing the length of the target fragments by design, they were able to capture degraded DNA fragments more efficiently. Furthermore, these authors adopted a "dual strand capture strategy" in which each of the two strands of a DNA duplex is captured independently. This approach enables comparison of the sequence obtained from both strands in a single double-stranded DNA molecule, facilitating differentiation of formalin-induced errors from true somatic mutations. This is particularly useful for low-frequency mutation detection in clinical cancer settings, which cannot be achieved using standard PCR-based assays.

10.7.2 Targeted Sequencing Using Gap-Fill Padlock Probes

In a related approach, gap-fill padlock probes were used with the ends hybridized a distance apart, spanning an entire exon (Porreca et al., 2007). A polymerase was then used to extend the spanned region and a ligase joined the extended sequence with the far end of the probe (Fig. 10.2C). As with the selector probes, a set of circular DNA molecules, all containing the sequence of interest along with a general synthetic sequence, was generated. Again, this could be amplified in multiplex, in this case using RCA followed by PCR and finally hRCA for the final library construction. An impressive 55,000 probes were deployed to target 6.7 megabase pairs of protein-coding sequences in the genome. The probes were synthesized through enzymatic amplification of a probe library synthesized on a microarray (Tian et al., 2004). The product was then analyzed on the Illumina Genome Analyzer; results showed that about 10,000 of the 55,000 targeted sequences were amplified and sequenced at least once. The approach has been developed further and applied to a global analysis of gene methylation signatures (Ball et al., 2009; Li et al., 2009b), allele-specific expression (Lee et al., 2009; Zhang et al., 2009), RNA editing (Li et al., 2009a), and genetic carrier screening (Umbarger et al., 2014).

10.8 IN SITU NUCLEIC ACID DETECTION USING PADLOCK PROBES

Early on, padlock probes were recognized as being useful for the analysis of DNA sequences in situ owing to the strong link formed between the circularized probe and its target strand (Nilsson et al., 1994). This opened up possibilities of observing genetic heterogeneity within cell populations and studying the localization of target molecules within individual cells. It was also shown that robust genotyping of repeated centromeric sequences differing in a single nucleotide position could be done on human metaphase chromosomes, revealing an unexpected distribution of these sequence variants among human chromosomes 13 and 21 (Nilsson et al., 1997). The technique was further developed using padlock probes synthesized by PCR to enable the synthesis of longer and thus more strongly labeled probes, and the number of centromeric repeats was extended to include chromosomes 7, 12, and 15 (Antson et al., 2000, 2003). In these studies, the probes were directly labeled with haptens or fluorophores. With this approach, however, it was not feasible to detect single DNA target molecules because it was impossible to distinguish specifically bound probe molecules from probes nonspecifically adsorbed to cells and the surface of the microscopy slide. This highlighted the need for a signal amplification scheme that strictly depended on the molecular recognition reaction.

10.8.1 Single-Molecule Detection and Genotyping In Situ

Similar to the single-molecule detection approach with RCA described in Section 10.4.1, individual RCPs can be generated and visualized in situ in cell and tissue samples in a strictly circularization-dependent reaction. This was first demonstrated by Lizardi et al. (1998) with the detection of nuclear sequences in cytologic halo preparations. There are indications that the RCA reaction is slowed down when it occurs under topologic inhibition, such as when the padlock probe is not able to free itself from the target strand (Banér et al., 1998). In response to this, Christian et al. (2001) presented an in situ target denaturation scheme employing the enzymatic preparation of ssDNA at the location of the target sequence. In this method, DNA is digested 5′ upstream of the target sequence. This is followed by removal of the nontarget strand using an exonuclease, leaving the target sequence single stranded and available for hybridization. The researchers combined this denaturation with padlock probe detection of the target sequence followed by RCA. Because the enzymatic denaturation method created a nearby free DNA end, the DNA polymerase was able to push off the padlock probe from the target strand when the topologic constraint became too high. The drawback with the presented denaturation procedure was that as the padlock probe falls off the target strand, the link between the target sequence and the RCA product is lost, increasing the risk of losing the signal from the slide or creating false-positive signals caused by drifting of RCA products into nearby cells. A variant of this denaturation process was later described in a robust protocol for SNP genotyping of mitochondrial DNA in situ (Larsson et al., 2004) (Figs. 10.6A and 10.7). Here, instead of digesting the DNA upstream of the target sequence, the cut takes place downstream of the target. This means that after exonucleolysis to make the target single-stranded, and padlock circularization on the target sequence, the nearby 3′-end of the target strand itself can be used to prime the following RCA reaction. In this way, not only is the topologic inhibition avoided, a covalent link is created between the target molecule and the positive detection signal that minimizes the risk of signal misplacement.

Larsson et al. (2004) also showed that this method can be used for relative target quantification. By counting differentially labeled RCA products in individual cells, the proportion of mutant versus wild-type mitochondrial DNA (mtDNA) molecules, differing in a single nucleotide could be determined. This ability to quantify different target molecules was further demonstrated in a more automated way, employing image analysis tools to perform digital counting of RCA products from detected mtDNA molecules in cultured cells (Jahangir Tafrechi et al., 2007). Dedicated software for RCA product identification and counting in digital images in a semiautomated fashion, BlobFinder, is freely available for users (Allalou and Wählby, http://www.cb.uu.se/~amin/BlobFinder).

10.8.2 Nuclear DNA Detection With Padlock Probes In Situ

Following early reports that padlock probes and RCA in situ could be used to detect specific sequences in genomic DNA and interphase nuclei (Christian et al., 2001; Lizardi et al., 1998), further demonstrations of the technique applied for this purpose were not published for a considerable time. This was probably because the efficiency of target detection with this technique was low, which made it difficult for regular researchers to apply it to their targets of interest. In fact, even when using the target-primed RCA for amplification of padlock probes, detection of mtDNA targets was estimated at about 10% of available targets (Larsson et al., 2004). Considering that nuclear DNA exhibits much tighter packing and more complex organization than mtDNA, it is not difficult to imagine the difficulties with finding proper assay conditions for detecting genomic DNA in low copy numbers. Instead, nuclear target sequences of high copy number have been successfully detected using target-primed RCA or similar techniques, providing useful steps toward the ultimate goal of efficient detection of nuclear single-copy sequences. Li and Young (2005) presented a proof-of-principle assay to detect double-strand breaks in mammalian cells using RCA primed by the target strand. In this method, double-strand breaks are introduced in cells by co-transfecting them with yeast HO endonuclease and its recognition site. When induced, HO endonuclease will perform digestion at integrated recognition sites, and site-specific cleavage is then detected by in situ RCA of padlock probes directed against the known sequence next to the double-strand breaks. The RCA assay was also combined with immunofluorescence, demonstrating the potential to perform studies of signaling pathways triggered by DNA double-strand breaks using this method.

Padlock probes and target-primed RCA in situ were also used in the comet assay to measure DNA damage and repair (Shaposhnikov et al., 2006). In the comet assay, cells are cast in agarose gels on objective slides, which are then subjected to gel electrophoresis. This causes cell nuclei to adopt a comet-like shape, with the amount of DNA forming the tail of the comet, reflecting the amount of DNA damage in the nucleus. In situ RCA is a preferred alternative to fluorescence in situ hybridization (FISH) in these types of cell preparations, because all steps can be executed at low temperatures and the background is low since signals are generated in a strictly target-dependent manner. For target detection with padlock probes in these preparations, the

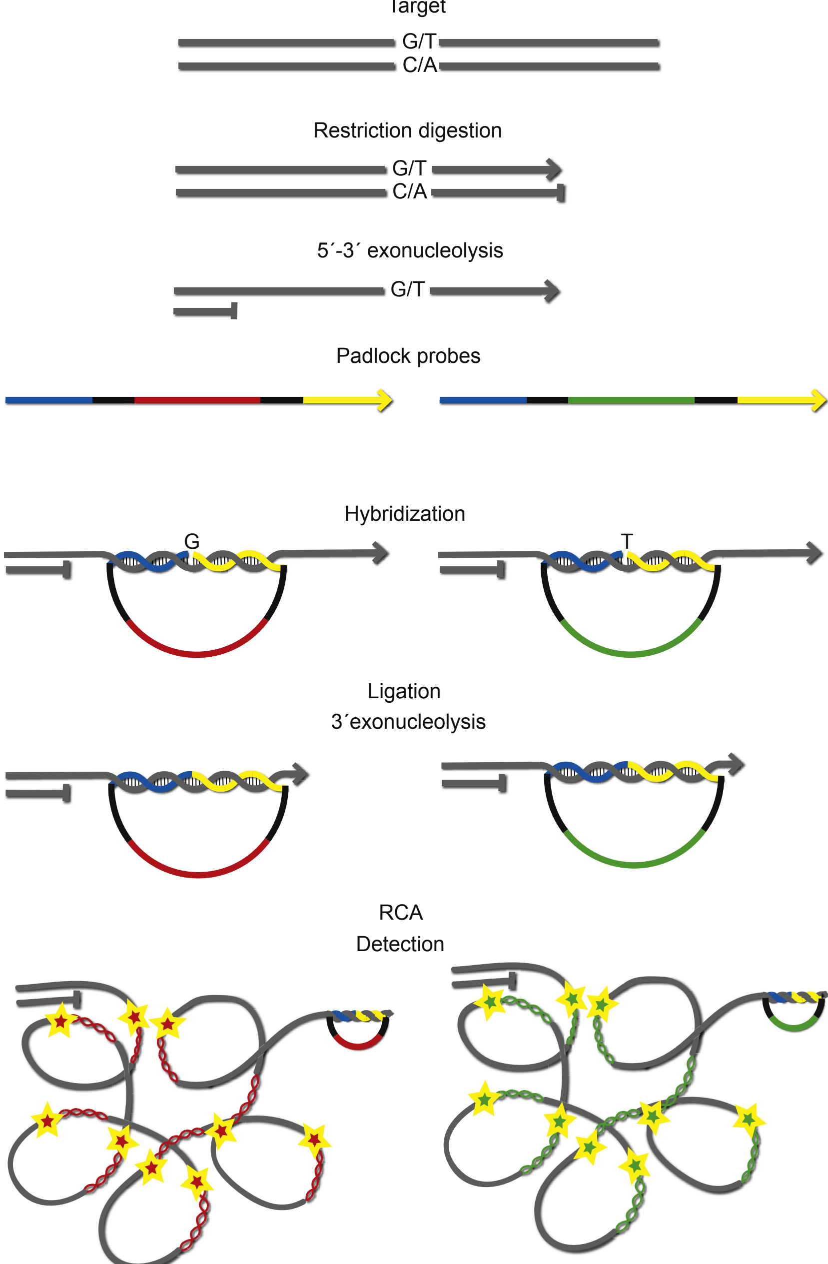


FIGURE 10.6 In situ genotyping using padlock probes and target-primed rolling circle amplification (RCA). The first steps of the procedure serve to prepare the target strands for probing and priming through a series of enzymatic treatments of fixed and permeabilized cells. First the target strands are digested at or 3′ to the target-sequence, to generate an end close to the padlock probe binding site. Then the nontarget strand is removed by 5′-exonucleolysis, making the target sequence single-stranded and available for hybridization. A pair of allele-specific padlock probes is added with target-complementary end-sequences (blue and yellow) that are identical except for the 3′-position where the single-nucleotide polymorphism is queried. The probes are equipped with two tag-sequences (red and green) to differentiate the amplification products from the two allele-specific padlock probes in the final tag-hybridization and fluorescence detection steps. After a combined hybridization and ligation step, the probes are rolling-circle amplified. The DNA polymerase used has potent 3′-exonucleolytic activity that degrades any 3′-protruding end of the target strand, creating a target strand that primes RCA of the padlock probe. A rolling-circle product (RCP) is created as an extension of the target strand, ensuring that the signal will remain localized with the target strand. The products from the two allele-specific padlock probes are differentially labeled with two fluorescence-tagged detection probes (red and green with yellow stars), hybridizing to the tag-sequence motif in the nontarget complementary part of the RCPs.

initial approach for target-primed RCA, with restriction digestion and exonucleolysis, was used to denature DNA to allow for detection of the nuclear 26−base pair (bp) Alu core sequence with sequence-specific probes and RCA (Fig. 10.7). This sequence is present in high numbers in each nucleus, as reflected in the amount of signal obtained

(A)

(B)

FIGURE 10.7 Detection of DNA in situ. The pictures show padlock probe and rolling circle amplification–based detection of different DNA sequences in situ. Scale bars represent 10 µm and genomic DNA is stained blue using 4′,6-diamidino-2-phenylindole. (A) Genotyping of the A3243G point mutation in human mitochondrial DNA (mtDNA) following the scheme outlined in Fig. 10.6. Two human cell lines were grown and fixed on a microscopy slide, one harboring the wild-type sequence of mtDNA and the other close to 100% of the mutant version. A pair of allele-specific padlock probes was designed such that the mutant probe generates green rolling-circle product (RCPs) in the detection step and the wild-type probe red RCPs. (B) Detection of the 26–base pair core Alu-repeat sequence in a comet preparation. In the comet assay, cells are immobilized on an agarose gel and subjected to alkali treatment to release the chromatin from the cell nucleus. Then an electric field is applied, which will primarily move damaged DNA to the tail of the comet while intact chromatin remains in the head of the comet. The head is to the right in the picture and the tail extends to the left. The RCPs were labeled with green fluorescence.

in genomic DNA with the in situ assay. Also, probes directed against mtDNA sequences were used, providing evidence that mtDNA disperses from the gels during the preparation of the comets.

Standard assays for in situ detection of nuclear DNA sequences, such as FISH, are commonly executed on condensed chromosomes in metaphase spreads to be able to identify and assign targets positively to specific chromosome locations. Chromosomes in metaphase are even more condensed than interphase nuclei; not surprisingly, early unsuccessful attempts to use in situ RCA to detect single-copy target sequences on metaphase chromosomes were reported (Lizardi et al., 1998; Zhong et al., 2001). Using the target primed approach, Lohmann et al. (2007)

demonstrated detection of repeated sequences on chromosomes Y and 6 on metaphase spreads. The detected targets were present in about 2000 and 12−51 copies per chromosome, respectively. These authors reported labeling of all targeted Y chromosomes at an efficiency of 1−10% of available targets, whereas detection of the less repeated target on chromosome 6 was less successful, with about 30% of spreads missing RCA products on one or two sister chromatids.

10.8.3 Padlock Probes for Detection of RNA In Situ

There is considerable interest in using padlock probes and RCA in situ to perform expression profiling and transcript genotyping in situ. One obstacle for this has long been decreased padlock probe ligation efficiency experienced when performing ligation on an RNA template (Nilsson et al., 2001, 2000). When considering the target-primed RCA, there has also been the question of whether polymerases are able to initiate DNA synthesis using an RNA primer.

Stougaard et al. (2007) presented a way to circumvent the problems of padlock probe ligation on RNA molecules by introducing a new probe format, turtle probes, for the detection of nonpolyadenylated RNA molecules in cultured cells. Turtle probes are similar to padlock probes in that they are also linear probes that are converted into circular molecules by ligating the two probe ends. In contrast to padlock probes, though, ligation does not take place using the target strand as template. Instead, the turtle probe folds part of itself into a hairpin structure in which an internal sequence of the probe serves as a template for joining the probe ends. In this study, 5S rRNA, Epstein−Barr virus-encoded RNA, and the RNA template for human telomerase were detected in FFPE cells and tissue. The presented method answers the question of whether single-stranded RNA can prime RCA, but it is still limited to abundant nonpolyadenylated RNA targets in which a target site can be chosen close to the 3′-end of the RNA. A more general approach for detecting RNA targets would thus be appreciated to be able to gain the full advantages of using short amplified oligonucleotide probes compared with traditional FISH methods.

Such an approach has been developed by Larsson et al. (2010) using padlock probes to target individual messenger RNA (mRNA) molecules directly in situ. mRNA molecules were first reverse-transcribed to cDNA and made single-stranded by the use of RNase H. After hybridization and ligation of padlock probes to the target cDNA molecule, ligated circles are locally amplified by RCA. These RCPs can be fluorescently labeled and result in diffraction limited spots that can be detected and analyzed in a fluorescent microscope (Larsson et al., 2010). Thanks to the ability of ligases to discriminate single-nucleotide mismatches, this approach can be used to detect disease-relevant mutations directly in cells or tissue. Grundberg et al. (2013) demonstrated somatic mutation detection in intratumor tissue targeting two *KRAS* codons, *EGFR* and *TP53* genes. Mutational status and expression profiles gained from such analyses can be used to predict tumor progression and response to treatment. The same approach was used to detect fusion transcripts in prostate cancer in fresh-frozen and FFPE samples (Kiflemariam et al., 2014).

The multiplex capability of such an assay is limited by the number of fluorophores that can be spectrally distinguished. To overcome this limitation, padlock probes can be bar-coded and in situ sequenced, as described in the following section.

10.8.4 In Situ Sequencing of Amplified Padlock Probes for High-Multiplexing Assays

A breakthrough in the application of a padlock probe-based identification assay was reported by Ke et al. (2013). In their work, they demonstrated the possibility of identifying different types of transcription profiles directly in single cells from fixed samples and preserved tissue with spatial resolution. This was achieved by performing a targeted sequencing approach in which a 4-bp target-specific barcode is introduced in the padlock probe that is sequenced using combinatorial probe-anchor ligation chemistry after subsequent circularization and RCA amplification (Drmanac et al., 2010). The method was applied to human fresh-frozen breast cancer tissue targeting 39 different mRNA transcripts simultaneously, and thus enabled localization-related gene expression analysis. The resulting expression profiles were compared with RNA-sequencing data finding a good correlation. In addition, by using gap-fill padlock probes it was possible to sequence five bases of the transcripts of interest. This was demonstrated by discriminating single-nucleotide variants in the actin, beta mRNA in human and mouse cells and by sequencing five bases in the KRAS gene detecting mutations in codons 12 and 13.

Lee et al. (2014) extended the multiplex ability and capacity of this approach by developing a method that allowed sequencing of up to 30 base reads using Sequencing by Oligonucleotide Ligation and Detection (SOLiD™) chemistry and different image analysis. To construct the libraries, RNA molecules are first reverse-transcribed to cDNA with random tagged hexamers and then circularized, amplified, and cross-linked to stabilize the RCPs. Subsequently, multiple cycles of sequencing by SOLiD ligation chemistry are performed (Porreca et al., 2006; Shendure et al., 2005). Lee et al. (2014) sequenced

human fibroblasts and identified 8102 annotated genes, allowing the identification of localized transcription profiles in individual cells during wound healing. This approach promises a new era in transcription profile analysis, because it opens up the possibility of analyzing localized RNA profiles that will provide valuable information about cell function and behavior in complex samples.

10.8.5 Padlock Probes for DNA Detection in Microbes

Interest has arisen in using padlock probes to detect DNA sequences in bacterial cells. Maruyama et al. (2005, 2006) first demonstrated how padlock probes could be used to visualize and count bacteria that had taken up free DNA from the environment in the form of plasmids. These authors also detected a single-copy gene of the bacterial genome, demonstrating the possibilities of using the technique to detect targets of lower abundance. The padlock probe technology is attractive for this type of work because relatively short (<40 nt) DNA sequences are targeted whereas cells that are unavailable for analysis using standard methods, such as selective cultivation, can be analyzed.

Smolina et al. (2007, 2008) used padlock probes to detect specific bacteria by targeting single-copy signature sites of different bacterial species. They used peptide nucleic acid (PNA) openers to denature double-stranded DNA locally and allow for padlock probe circularization. This approach avoids global DNA denaturation, which could be advantageous for some applications but also limits detection to specific sites in the genome, hence limiting the general applicability of this method.

Wamsley and Barbet (2008) used endonuclease digestion instead of PNA openers to make the bacterial genome linear and successfully detected *Anaplasma* spp. in infected endothelial and myeloblastic leukemia cells.

10.9 AUTOMATION AND MINIATURIZATION OF PADLOCK PROBE/ROLLING CIRCLE AMPLIFICATION ASSAYS

As described in previous sections, the versatility and advantages of DNA circularization-based detection assays are numerous, and several formats have been developed that demonstrate the promise of these molecular tools for medical applications. Nevertheless, these assays often necessitate multiple-step protocols that require the handling of different enzymes and reagents, not to mention the expertise needed to manipulate such components and to read and interpret the final results, which therefore relegates the use of these methods purely to specialized research laboratories. Microfluidics is a technology that enables the manipulation of liquids in micron-size structures (see also Chapter 11). Over the past decade, this technology has emerged as the preferred toolbox to achieve the automation and miniaturization of assays (lab-on-a chip), offering multiple advantages such as small volume consumption, automation, integration of multiple steps, and reduced manufacturing cost for scalable devices (Haeberle and Zengerle, 2007; Lei, 2012; Thorsen et al., 2002).

Some effort has been undertaken to automate the assays and applications described in this chapter. The first approach was taken by Mahmoudian et al. (2008), who implemented in a single polymer microchip both RCA and C2CA reactions combined with electrophoretic analysis. They reported fivefold lower reagent consumption and a detection limit of 25 ng of *Vibrio cholerae* genomic DNA in less than 65 min. Sato et al. (2010) developed a microfluidic channel filled with sepharose beads that was used to capture and detect DNA with padlock probes and RCA. The channel was designed in a way that facilitated imaging and quantification of RCPs directly onto the surface of the beads, thereby integrating sample processing, amplification, and detection. Similarly, a glass microchip with a functionalized microchannel surface was developed to manipulate small volume samples. A special surface chemistry was developed to graft specific DNA oligonucleotides, in which target molecules were captured and subjected to padlock probe detection and RCA. The channel was designed to be 100×42 μm and was proven able to process samples as small as 5 nL (Tanaka et al., 2011).

Digital microfluidics, a technology that uses electrowetting, allows the precise transport of small volumes by manipulating the hydrophilicity of a surface upon applying an electric field (Abdelgawad and Wheeler, 2009; Jebrail et al., 2012). Using this technology, a semiautomated microchip for performing magnetic bead–based C2CA with high performance was developed (Kühnemund et al., 2014). The same results were obtained as those in a manually performed assay in PCR tubes, and the researchers reported a sensitivity of 1 aM with a wide dynamic range. In addition, no compartments were needed to perform the different reactions. This provided an efficient way to mix and handle magnetic beads, which has been troublesome for other kinds of microfluidic platforms.

An automated microfluidic system has also been developed to perform the in situ detection of RNA by padlock probe/RCA described in Section 10.8.3. This system showed for the first time the complete automated handling of reagents, flow, and temperature control displaying good performance compared with the manual procedure (Kuroda et al., 2014).

Currently, some developed platforms are close to commercialization. For instance, the concept described in Section 10.4.1 for single-molecule counting has been integrated and automated into a dedicated instrument that automatically counts single RCPs in less than 3 min per sample (Q-linea AB; Aquila 400). The same company developed a second-generation fluidic instrument that can collect, process, and analyze air samples for the detection of pathogens using a magnetic bead−based C2CA (Aquila 1000). A full-scale test of this prototype has been successfully tested in a real-life scenario (www.qlinea.com).

10.10 CONCLUSIONS

As described in this chapter, intramolecular DNA circularization assays have found various important applications in molecular medicine. There is likely more to come because some of the techniques are yet not mature but are still promising. The targeted multiplexed genome amplification approaches described in Section 10.7, for example, have become a standard library-preparation step for high-throughput DNA sequencers, particularly in applications where the source of DNA is limiting, such as in tumor biopsies, and where really deep sequence coverage is desired.

Circularization techniques combined with RCA have some unique and promising properties for rapid diagnostics, such as those outlined in Section 10.5. Application areas may include infectious diagnostics and high-performance multiplexed quantitative expression analyses. These assays will be highly favored by automated solutions that can potentially be made available by developing devices and/or platforms that can be placed onsite or even for point-of-care configurations. Finally, the single-molecule in situ genotyping technique described in Section 10.8 is unique and may find a broad application in biomedical research and diagnostics. One example of a relevant clinical pathology investigation is to find rare surviving malignant cells in tissue biopsies taken after therapy. Because the approach is multiplexible, it could also be applied in cell differentiation studies to determine the colocalization of several lineage-specific transcripts in developing embryos or organs, including tumor tissue, thereby tracking the developmental history of individual cells in complex tissues.

There is the potential for many exciting developments of DNA circularization assays in addition to the ones mentioned here. All of the possibilities that come with this approach, such as inherent multiplexibility, single-molecule sensitivity, digital quantification, and localized detection, are being explored and developed.

REFERENCES

Abdelgawad, M., Wheeler, A.R., 2009. The digital revolution: a new paradigm for microfluidics. Adv. Mater. 21, 920−925. http://dx.doi.org/10.1002/adma.200802244.

Absalan, F., Ronaghi, M., 2007. Molecular inversion probe assay. Methods Mol. Biol. http://dx.doi.org/10.1007/978-1-59745-515-2_20.

Ahmed, S.A., van den Ende, B.H.G.G., Fahal, A.H., van de Sande, W.W.J., de Hoog, G.S., 2014. Rapid identification of black grain eumycetoma causative agents using rolling circle amplification. PLoS Negl. Trop. Dis. 8, e3368. http://dx.doi.org/10.1371/journal.pntd.0003368.

Akhras, M.S., Unemo, M., Thiyagarajan, S., Nyrén, P., Davis, R.W., Fire, A.Z., Pourmand, N., 2007. Connector inversion probe technology: a powerful one-primer multiplex DAN amplification system for numerous scientific application. PLoS One 2, e915. http://dx.doi.org/10.1371/journal.pone.0000915.

Albert, T.J., Molla, M.N., Muzny, D.M., Nazareth, L., Wheeler, D., Song, X., Richmond, T.A., Middle, C.M., Rodesch, M.J., Packard, C.J., Weinstock, G.M., Gibbs, R.A., 2007. Direct selection of human genomic loci by microarray hybridization. Nat. Methods 4, 903−905. http://dx.doi.org/10.1038/nmeth1111.

Alsmadi, O.A., Bornarth, C.J., Song, W., Wisniewski, M., Du, J., Brockman, J.P., Faruqi, A.F., Hosono, S., Sun, Z., Du, Y., Wu, X., Egholm, M., Abarzúa, P., Lasken, R.S., Driscoll, M.D., 2003. High accuracy genotyping directly from genomic DNA using a rolling circle amplification based assay. BMC Genomics 4, 21. http://dx.doi.org/10.1186/1471-2164-4-21.

Antson, D.O., Isaksson, A., Landegren, U., Nilsson, M., 2000. PCR-generated padlock probes detect single nucleotide variation in genomic DNA. Nucleic Acids Res. 28, E58. http://dx.doi.org/10.1093/nar/28.12.e58.

Antson, D.-O., Mendel-Hartvig, M., Landegren, U., Nilsson, M., 2003. PCR-generated padlock probes distinguish homologous chromosomes through quantitative fluorescence analysis. Eur. J. Hum. Genet. 11, 357−363. http://dx.doi.org/10.1038/sj.ejhg.5200966.

Ball, M.P., Li, J.B., Gao, Y., Lee, J.-H., LeProust, E.M., Park, I.-H., Xie, B., Daley, G.Q., Church, G.M., 2009. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat. Biotechnol. 27, 361−368. http://dx.doi.org/10.1038/nbt.1533.

Banér, J., Gyarmati, P., Yacoub, A., Hakhverdyan, M., Stenberg, J., Ericsson, O., Nilsson, M., Landegren, U., Belák, S., 2007. Microarray-based molecular detection of foot-and-mouth disease, vesicular stomatitis and swine vesicular disease viruses, using padlock probes. J. Virol. Methods 143, 200−206. http://dx.doi.org/10.1016/j.jviromet.2007.03.004.

Banér, J., Isaksson, A., Waldenström, E., Jarvius, J., Landegren, U., Nilsson, M., 2003. Parallel gene analysis with allele-specific padlock probes and tag microarrays. Nucleic Acids Res. 31, e103. http://dx.doi.org/10.1093/nar/gng104.

Banér, J., Marits, P., Nilsson, M., Winqvist, O., Landegren, U., 2005. Analysis of T-cell receptor Vβ gene repertoires after immune stimulation and in malignancy by use of padlock probes and microarrays. Clin. Chem. 51, 768−775. http://dx.doi.org/10.1373/clinchem.2004.047266.

Banér, J., Nilsson, M., Mendel-Hartvig, M., Landegren, U., 1998. Signal amplification of padlock probes by rolling circle replication. Nucleic Acids Res. 26, 5073−5078. http://dx.doi.org/10.1093/nar/26.22.5073.

Bejhed, R.S., de la Torre, T.Z.G., Donolato, M., Hansen, M.F., Svedlindh, P., Strömberg, M., 2015. Turn-on optomagnetic bacterial DNA sequence detection using volume-amplified magnetic nanobeads. Biosens. Bioelectron. 66, 405–411. http://dx.doi.org/10.1016/j.bios.2014.11.048.

Blab, G.A., Schmidt, T., Nilsson, M., 2004. Homogeneous detection of single rolling circle replication products. Anal. Chem. 76, 495–498. http://dx.doi.org/10.1021/ac034987.

Christian, A.T., Pattee, M.S., Attix, C.M., Reed, B.E., Sorensen, K.J., Tucker, J.D., 2001. Detection of DNA point mutations and mRNA expression levels by rolling circle amplification in individual cells. Proc. Natl. Acad. Sci. U.S.A. 98, 14238–14243. http://dx.doi.org/10.1073/pnas.251383598.

Consortium, T.I.H., 2005. A haplotype map of the human genome. Nature 437, 1299–1320. http://dx.doi.org/10.1038/nature04226.

Cuyvers, E., De Roeck, A., Van den Bossche, T., Van Cauwenberghe, C., Bettens, K., Vermeulen, S., Mattheijssens, M., Peeters, K., Engelborghs, S., Vandenbulcke, M., Vandenberghe, R., De Deyn, P.P., Van Broeckhoven, C., Sleegers, K., 2015. Mutations in ABCA7 in a Belgian cohort of Alzheimer's disease patients: a targeted resequencing study. Lancet Neurol. 14, 814–822. http://dx.doi.org/10.1016/S1474-4422(15)00133-7.

Dahl, F., Banér, J., Gullberg, M., Mendel-Hartvig, M., Landegren, U., Nilsson, M., 2004. Circle-to-circle amplification for precise and sensitive DNA analysis. Proc. Natl. Acad. Sci. U.S.A. 101, 4548–4553. http://dx.doi.org/10.1073/pnas.0400834101.

Dahl, F., Stenberg, J., Fredriksson, S., Welch, K., Zhang, M., Nilsson, M., Bicknell, D., Bodmer, W.F., Davis, R.W., Ji, H., 2007. Multigene amplification and massively parallel sequencing for cancer mutation discovery. Proc. Natl. Acad. Sci. U.S.A. 104, 9387–9392. http://dx.doi.org/10.1073/pnas.0702165104.

de la Torre, T.Z.G., Mezger, A., Herthnek, D., Johansson, C., Svedlindh, P., Nilsson, M., Strømme, M., 2011. Detection of rolling circle amplified DNA molecules using probe-tagged magnetic nanobeads in a portable AC susceptometer. Biosens. Bioelectron. 29, 195–199. http://dx.doi.org/10.1016/j.bios.2011.08.019.

Dean, F.B., Hosono, S., Fang, L., Wu, X., Faruqi, A.F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., Driscoll, M., Song, W., Kingsmore, S.F., Egholm, M., Lasken, R.S., 2002. Comprehensive human genome amplification using multiple displacement amplification. Proc. Natl. Acad. Sci. U.S.A. 99, 5261–5266. http://dx.doi.org/10.1073/pnas.082089499.

Donolato, M., Antunes, P., Bejhed, R.S., Zardán Gómez de la Torre, T., Østerberg, F.W., Strömberg, M., Nilsson, M., Strømme, M., Svedlindh, P., Hansen, M.F., Vavassori, P., 2015. Novel readout method for molecular diagnostic assays based on optical measurements of magnetic nanobead dynamics. Anal. Chem. 87, 1622–1629. http://dx.doi.org/10.1021/ac503191v.

Dressman, D., Yan, H., Traverso, G., Kinzler, K.W., Vogelstein, B., 2003. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. Proc. Natl. Acad. Sci. U.S.A. 100, 8817–8822. http://dx.doi.org/10.1073/pnas.1133470100.

Drmanac, R., Sparks, A.B., Callow, M.J., Halpern, A.L., Burns, N.L., Kermani, B.G., Carnevali, P., Nazarenko, I., Nilsen, G.B., Yeung, G., Dahl, F., Fernandez, A., Staker, B., Pant, K.P., Baccash, J., Borcherding, A.P., Brownley, A., Cedeno, R., Chen, L., Chernikoff, D., Cheung, A., Chirita, R., Curson, B., Ebert, J.C., Hacker, C.R., Hartlage, R., Hauser, B., Huang, S., Jiang, Y., Karpinchyk, V., Koenig, M., Kong, C., Landers, T., Le, C., Liu, J., McBride, C.E., Morenzoni, M., Morey, R.E., Mutch, K., Perazich, H., Perry, K., Peters, B.A., Peterson, J., Pethiyagoda, C.L., Pothuraju, K., Richter, C., Rosenbaum, A.M., Roy, S., Shafto, J., Sharanhovich, U., Shannon, K.W., Sheppy, C.G., Sun, M., Thakuria, J.V., Tran, A., Vu, D., Zaranek, A.W., Wu, X., Drmanac, S., Oliphant, A.R., Banyai, W.C., Martin, B., Ballinger, D.G., Church, G.M., Reid, C.A., 2010. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science 327, 78–81. http://dx.doi.org/10.1126/science.1181498.

Eichler, E.E., Nickerson, D.A., Altshuler, D., Bowcock, A.M., Brooks, L.D., Carter, N.P., Church, D.M., Felsenfeld, A., Guyer, M., Lee, C., Lupski, J.R., Mullikin, J.C., Pritchard, J.K., Sebat, J., Sherry, S.T., Smith, D., Valle, D., Waterston, R.H., 2007. Completing the map of human genetic variation. Nature 447, 161–165. http://dx.doi.org/10.1038/447161a.

Engström, A., Zardán Gómez de la Torre, T., Strømme, M., Nilsson, M., Herthnek, D., 2013. Detection of rifampicin resistance in *Mycobacterium tuberculosis* by padlock probes and magnetic nanobead-based readout. PLoS One 8, e62015. http://dx.doi.org/10.1371/journal.pone.0062015.

Ericsson, O., Jarvius, J., Schallmeiner, E., Howell, M., Nong, R.Y., Reuter, H., Hahn, M., Stenberg, J., Nilsson, M., Landegren, U., 2008. A dual-tag microarray platform for high-performance nucleic acid and protein analyses. Nucleic Acids Res. 36, e45. http://dx.doi.org/10.1093/nar/gkn106.

Eriksson, R., Jobs, M., Ekstrand, C., Ullberg, M., Herrmann, B., Landegren, U., Nilsson, M., Blomberg, J., 2009. Multiplex and quantifiable detection of nucleic acid from pathogenic fungi using padlock probes, generic real time PCR and specific suspension array readout. J. Microbiol. Methods 78, 195–202. http://dx.doi.org/10.1016/j.mimet.2009.05.016.

Faruqi, A.F., Hosono, S., Driscoll, M.D., Dean, F.B., Alsmadi, O., Bandaru, R., Kumar, G., Grimwade, B., Zong, Q., Sun, Z., Du, Y., Kingsmore, S., Knott, T., Lasken, R.S., 2001. High-throughput genotyping of single nucleotide polymorphisms with rolling circle amplification. BMC Genomics 2, 4. http://dx.doi.org/10.1186/1471-2164-2-4.

Feng, P., Klaassen, C.H.W., Meis, J.F., Najafzadeh, M.J., Gerrits Van Den Ende, A.H.G., Xi, L., de Hoog, G.S., 2013. Identification and typing of isolates of Cyphellophora and relatives by use of amplified fragment length polymorphism and rolling circle amplification. J. Clin. Microbiol. 51, 931–937. http://dx.doi.org/10.1128/JCM.02898-12.

Fire, A., Xu, S.Q., 1995. Rolling replication of short DNA circles. Proc. Natl. Acad. Sci. U.S.A. 92, 4641–4645. http://dx.doi.org/10.1073/pnas.92.10.4641.

Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., Pasternak, S., Wheeler, D.A., Willis, T.D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu, J., Zhang, B., Zhang, Q., Zhao, H., Zhao, H., Zhou, J., Gabriel, S.B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R.C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z.,

Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Waye, M.M.Y., Tsui, S.K.W., Xue, H., Wong, J.T.-F., Galver, L.M., Fan, J.-B., Gunderson, K., Murray, S.S., Oliphant, A.R., Chee, M.S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.-F., Phillips, M.S., Roumy, S., Sallée, C., Verner, A., Hudson, T.J., Kwok, P.-Y., Cai, D., Koboldt, D.C., Miller, R.D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.-C., Mak, W., Song, Y.Q., Tam, P.K.H., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., Tanaka, T., Tsunoda, T., Deloukas, P., Bird, C.P., Delgado, M., Dermitzakis, E.T., Gwilliam, R., Hunt, S., Morrison, J., Powell, D., Stranger, B.E., Whittaker, P., Bentley, D.R., Daly, M.J., de Bakker, P.I.W., Barrett, J., Chretien, Y.R., Maller, J., McCarroll, S., Patterson, N., Pe'er, I., Price, A., Purcell, S., Richter, D.J., Sabeti, P., Saxena, R., Schaffner, S.F., Sham, P.C., Varilly, P., Altshuler, D., Stein, L.D., Krishnan, L., Smith, A.V., Tello-Ruiz, M.K., Thorisson, G.A., Chakravarti, A., Chen, P.E., Cutler, D.J., Kashuk, C.S., Lin, S., Abecasis, G.R., Guan, W., Li, Y., Munro, H.M., Qin, Z.S., Thomas, D.J., McVean, G., Auton, A., Bottolo, L., Cardin, N., Eyheramendy, S., Freeman, C., Marchini, J., Myers, S., Spencer, C., Stephens, M., Donnelly, P., Cardon, L.R., Clarke, G., Evans, D.M., Morris, A.P., Weir, B.S., Tsunoda, T., Mullikin, J.C., Sherry, S.T., Feolo, M., Skol, A., Zhang, H., Zeng, C., Zhao, H., Matsuda, I., Fukushima, Y., Macer, D.R., Suda, E., Rotimi, C.N., Adebamowo, C.A., Ajayi, I., Aniagwu, T., Marshall, P.A., Nkwodimmah, C., Royal, C.D.M., Leppert, M.F., Dixon, M., Peiffer, A., Qiu, R., Kent, A., Kato, K., Niikawa, N., Adewole, I.F., Knoppers, B.M., Foster, M.W., Clayton, E.W., Watkin, J., Gibbs, R.A., Belmont, J.W., Muzny, D., Nazareth, L., Sodergren, E., Weinstock, G.M., Wheeler, D.A., Yakub, I., Gabriel, S.B., Onofrio, R.C., Richter, D.J., Ziaugra, L., Birren, B.W., Daly, M.J., Altshuler, D., Wilson, R.K., Fulton, L.L., Rogers, J., Burton, J., Carter, N.P., Clee, C.M., Griffiths, M., Jones, M.C., McLay, K., Plumb, R.W., Ross, M.T., Sims, S.K., Willey, D.L., Chen, Z., Han, H., Kang, L., Godbout, M., Wallenburg, J.C., L'Archevêque, P., Bellemare, G., Saeki, K., Wang, H., An, D., Fu, H., Li, Q., Wang, Z., Wang, R., Holden, A.L., Brooks, L.D., McEwen, J.E., Guyer, M.S., Wang, V.O., Peterson, J.L., Shi, M., Spiegel, J., Sung, L.M., Zacharia, L.F., Collins, F.S., Kennedy, K., Jamieson, R., Stewart, J., 2007. A second generation human haplotype map of over 3.1 million SNPs. Nature 449, 851−861. http://dx.doi.org/10.1038/nature06258.

Gibbs, R.A., Belmont, J.W., Hardenbol, P., Willis, T.D., Yu, F., Yang, H., Ch'ang, L.-Y., Huang, W., Liu, B., Shen, Y., Tam, P.K.-H., Tsui, L.-C., Waye, M.M.Y., Wong, J.T.-F., Zeng, C., Zhang, Q., Chee, M.S., Galver, L.M., Kruglyak, S., Murray, S.S., Oliphant, A.R., Montpetit, A., Hudson, T.J., Chagnon, F., Ferretti, V., Leboeuf, M., Phillips, M.S., Verner, A., Kwok, P.-Y., Duan, S., Lind, D.L., Miller, R.D., Rice, J.P., Saccone, N.L., Taillon-Miller, P., Xiao, M., Nakamura, Y., Sekine, A., Sorimachi, K., Tanaka, T., Tanaka, Y., Tsunoda, T., Yoshino, E., Bentley, D.R., Deloukas, P., Hunt, S., Powell, D., Altshuler, D., Gabriel, S.B., Zhang, H., Zeng, C., Matsuda, I., Fukushima, Y., Macer, D.R., Suda, E., Rotimi, C.N., Adebamowo, C.A., Aniagwu, T., Marshall, P.A., Matthew, O., Nkwodimmah, C., Royal, C.D.M., Leppert, M.F., Dixon, M., Stein, L.D., Cunningham, F., Kanani, A., Thorisson, G.A., Chakravarti, A., Chen, P.E., Cutler, D.J., Kashuk, C.S., Donnelly, P., Marchini, J., McVean, G.A.T., Myers, S.R., Cardon, L.R., Abecasis, G.R., Morris, A., Weir, B.S., Mullikin, J.C., Sherry, S.T., Feolo, M., Altshuler, D., Daly, M.J., Schaffner, S.F., Qiu, R., Kent, A., Dunston, G.M., Kato, K., Niikawa, N., Knoppers, B.M., Foster, M.W., Clayton, E.W., Wang, V.O., Watkin, J., Gibbs, R.A., Belmont, J.W., Sodergren, E., Weinstock, G.M., Wilson, R.K., Fulton, L.L., Rogers, J., Birren, B.W., Han, H., Wang, H., Godbout, M., Wallenburg, J.C., L'Archevêque, P., Bellemare, G., Todani, K., Fujita, T., Tanaka, S., Holden, A.L., Lai, E.H., Collins, F.S., Brooks, L.D., McEwen, J.E., Guyer, M.S., Jordan, E., Peterson, J.L., Spiegel, J., Sung, L.M., Zacharia, L.F., Kennedy, K., Dunn, M.G., Seabrook, R., Shillito, M., Skene, B., Stewart, J.G., Valle (chair), D.L., Clayton (co-chair), E.W., Jorde (co-chair), L.B., Belmont, J.W., Chakravarti, A., Cho, M.K., Duster, T., Foster, M.W., Jasperse, M., Knoppers, B.M., Kwok, P.-Y., Licinio, J., Long, J.C., Marshall, P.A., Ossorio, P.N., Wang, V.O., Rotimi, C.N., Royal, C.D.M., Spallone, P., Terry, S.F., Lander (chair), E.S., Lai (co-chair), E.H., Nickerson (co-chair), D.A., Abecasis, G.R., Altshuler, D., Bentley, D.R., Boehnke, M., Cardon, L.R., Daly, M.J., Deloukas, P., Douglas, J.A., Gabriel, S.B., Hudson, R.R., Hudson, T.J., Kruglyak, L., Kwok, P.-Y., Nakamura, Y., Nussbaum, R.L., Royal, C.D.M., Schaffner, S.F., Sherry, S.T., Stein, L.D., Tanaka, T., 2003. The International HapMap project. Nature. http://dx.doi.org/10.1038/nature02168.

Göransson, J., Ke, R., Nong, R.Y., Howell, W.M., Karman, A., Grawé, J., Stenberg, J., Granberg, M., Elgh, M., Herthnek, D., Wikström, P., Jarvius, J., Nilsson, M., 2012. Rapid identification of bio-molecules applied for detection of biosecurity agents using rolling circle amplification. PLoS One 7, e31068. http://dx.doi.org/10.1371/journal.pone.0031068.

Göransson, J., Wählby, C., Isaksson, M., Howell, W.M., Jarvius, J., Nilsson, M., 2009. A single molecule array for digital targeted molecular analyses. Nucleic Acids Res. 37, e7. http://dx.doi.org/10.1093/nar/gkn921.

Grundberg, I., Kiflemariam, S., Mignardi, M., Imgenberg-Kreuz, J., Edlund, K., Micke, P., Sundström, M., Sjöblom, T., Botling, J., Nilsson, M., 2013. In situ mutation detection and visualization of intratumor heterogeneity for cancer research and diagnostics. Oncotarget 4, 2407−2418.

Gyarmati, P., Conze, T., Zohari, S., LeBlanc, N., Nilsson, M., Landegren, U., Banér, J., Belák, S., 2008. Simultaneous genotyping of all hemagglutinin and neuraminidase subtypes of avian influenza viruses by use of padlock probes. J. Clin. Microbiol. 46, 1747−1751. http://dx.doi.org/10.1128/JCM.02292-07.

Haeberle, S., Zengerle, R., 2007. Microfluidic platforms for lab-on-a-chip applications. Lab. Chip 7, 1094−1110. http://dx.doi.org/10.1039/b706364b.

Hardenbol, P., Banér, J., Jain, M., Nilsson, M., Namsaraev, E.A., Karlin-Neumann, G.A., Fakhrai-Rad, H., Ronaghi, M., Willis, T.D., Landegren, U., Davis, R.W., 2003. Multiplexed genotyping with sequence-tagged molecular inversion probes. Nat. Biotechnol. 21, 673−678. http://dx.doi.org/10.1038/nbt821.

Hardenbol, P., Yu, F., Belmont, J., MacKenzie, J., Bruckner, C., Brundage, T., Boudreau, A., Chow, S., Eberle, J., Erbilgin, A., Falkowski, M., Fitzgerald, R., Ghose, S., Iartchouk, O., Jain, M., Karlin-Neumann, G., Lu, X., Miao, X., Moore, B., Moorhead, M., Namsaraev, E., Pasternak, S., Prakash, E., Tran, K., Wang, Z., Jones, H.B., Davis, R.W., Willis, T.D., Gibbs, R.A., 2005. Highly

multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay. Genome Res. 15, 269–275. http://dx.doi.org/10.1101/gr.3185605.

Isaksson, M., Stenberg, J., Dahl, F., Thuresson, A.C., Bondeson, M.L., Nilsson, M., 2007. MLGA – a rapid and cost-efficient assay for gene copy-number analysis. Nucleic Acids Res. 35, e115. http://dx.doi.org/10.1093/nar/gkm651.

Jahangir Tafrechi, R.S., van de Rijke, F.M., Allallou, A., Larsson, C., Sloos, W.C.R., van de Sande, M., Wählby, C., Janssen, G.M.C., Raap, A.K., 2007. Single-cell A3243G mitochondrial DNA mutation load assays for segregation analysis. J. Histochem. Cytochem. 55, 1159–1166. http://dx.doi.org/10.1369/jhc.7A7282.2007.

Jarvius, J., Melin, J., Göransson, J., Stenberg, J., Fredriksson, S., Gonzalez-Rey, C., Bertilsson, S., Nilsson, M., 2006. Digital quantification using amplified single-molecule detection. Nat. Methods 3, 725–727. http://dx.doi.org/10.1038/nmeth916.

Jebrail, M.J., Bartsch, M.S., Patel, K.D., 2012. Digital microfluidics: a versatile tool for applications in chemistry, biology and medicine. Lab. Chip 12, 2452. http://dx.doi.org/10.1039/c2lc40318h.

Johansson, H., Isaksson, M., Sörqvist, E.F., Roos, F., Stenberg, J., Sjöblom, T., Botling, J., Micke, P., Edlund, K., Fredriksson, S., Kultima, H.G., Ericsson, O., Nilsson, M., 2011. Targeted resequencing of candidate genes using selector probes. Nucleic Acids Res. 39, e8. http://dx.doi.org/10.1093/nar/gkq1005.

Juul, S., Ho, Y.P., Koch, J., Andersen, F.F., Stougaard, M., Leong, K.W., Knudsen, B.R., 2011. Detection of single enzymatic events in rare or single cells using microfluidics. ACS Nano 5, 8305–8310. http://dx.doi.org/10.1021/nn203012q.

Ke, R., Mignardi, M., Pacureanu, A., Svedlund, J., Botling, J., Wählby, C., Nilsson, M., 2013. In situ sequencing for RNA analysis in preserved tissue and cells. Nat. Methods 10, 857–860. http://dx.doi.org/10.1038/nmeth.2563.

Ke, R., Zorzet, A., Göransson, J., Lindegren, G., Sharifi-Mood, B., Chinikar, S., Mardani, M., Mirazimi, A., Nilsson, M., 2011. Colorimetric nucleic acid testing assay for RNA virus detection based on circle-to-circle amplification of padlock probes. J. Clin. Microbiol. 49, 4279–4285. http://dx.doi.org/10.1128/JCM.00713-11.

Khatkar, M.S., Zenger, K.R., Hobbs, M., Hawken, R.J., Cavanagh, J.A.L., Barris, W., McClintock, A.E., McClintock, S., Thomson, P.C., Tier, B., Nicholas, F.W., Raadsma, H.W., 2007. A primary assembly of a bovine haplotype block map based on a 15,036-single-nucleotide polymorphism panel genotyped in Holstein-Friesian cattle. Genetics 176, 763–772. http://dx.doi.org/10.1534/genetics.106.069369.

Kidd, J.M., Cooper, G.M., Donahue, W.F., Hayden, H.S., Sampas, N., Graves, T., Hansen, N., Teague, B., Alkan, C., Antonacci, F., Haugen, E., Zerr, T., Yamada, N.A., Tsang, P., Newman, T.L., Tüzün, E., Cheng, Z., Ebling, H.M., Tusneem, N., David, R., Gillett, W., Phelps, K.A., Weaver, M., Saranga, D., Brand, A., Tao, W., Gustafson, E., McKernan, K., Chen, L., Malig, M., Smith, J.D., Korn, J.M., McCarroll, S.A., Altshuler, D.A., Peiffer, D.A., Dorschner, M., Stamatoyannopoulos, J., Schwartz, D., Nickerson, D.A., Mullikin, J.C., Wilson, R.K., Bruhn, L., Olson, M.V., Kaul, R., Smith, D.R., Eichler, E.E., 2008. Mapping and sequencing of structural variation from eight human genomes. Nature 453, 56–64. http://dx.doi.org/10.1038/nature06862.

Kiflemariam, S., Mignardi, M., Ali, M.A., Bergh, A., Nilsson, M., Sjöblom, T., 2014. In situ sequencing identifies TMPRSS2–ERG fusion transcripts, somatic point mutations and gene expression levels in prostate cancers. J. Pathol. 234, 253–261. http://dx.doi.org/10.1002/path.4392.

Kim, E.-H., Lee, S., Park, J., Lee, K., Bhak, J., Kim, B.C., 2014. New lung cancer panel for high-throughput targeted resequencing. Genomics Inform. 12, 50–57. http://dx.doi.org/10.5808/GI.2014.12.2.50.

Kong, F., Tong, Z., Chen, X., Sorrell, T., Wang, B., Wu, Q., Ellis, D., Chen, S., 2008. Rapid identification and differentiation of Trichophyton species, based on sequence polymorphisms of the ribosomal internal transcribed spacer regions, by rolling-circle amplification. J. Clin. Microbiol. 46, 1192–1199. http://dx.doi.org/10.1128/JCM.02235-07.

Konry, T., Smolina, I., Yarmush, J.M., Irimia, D., Yarmush, M.L., 2011. Ultrasensitive detection of low-abundance surface-marker protein using isothermal rolling circle amplification in a microfluidic nanoliter platform. Small 7, 395–400. http://dx.doi.org/10.1002/smll.201001620.

Kühnemund, M., Nilsson, M., 2014. Digital quantification of rolling circle amplified single DNA molecules in a resistive pulse sensing nanopore. Biosens. Bioelectron. 67, 1–7. http://dx.doi.org/10.1016/j.bios.2014.06.040.

Kühnemund, M., Witters, D., Nilsson, M., Lammertyn, J., 2014. Circle-to-circle amplification on a digital microfluidic chip for amplified single molecule detection. Lab. Chip 14, 2983–2992. http://dx.doi.org/10.1039/c4lc00348a.

Kuroda, A., Ishigaki, Y., Nilsson, M., Sato, K., Sato, K., 2014. Microfluidics-based in situ padlock/rolling circle amplification system for counting single DNA molecules in a cell. Anal. Sci. 30, 1107–1112. http://dx.doi.org/10.2116/analsci.30.1107.

Landegren, U., Kaiser, R., Sanders, J., Hood, L., 1988. A ligase-mediated gene detection technique. Science 241, 1077–1080. http://dx.doi.org/10.1126/science.3413476.

Larsson, C., Grundberg, I., Soderberg, O., Nilsson, M., 2010. In situ detection and genotyping of individual mRNA molecules. Nat. Methods 7, 395–397.

Larsson, C., Koch, J., Nygren, A., Janssen, G., Raap, A.K., Landegren, U., Nilsson, M., 2004. In situ genotyping individual DNA molecules by target-primed rolling-circle amplification of padlock probes. Nat. Methods 1, 227–232. http://dx.doi.org/10.1038/nmeth723.

Laurie, C.C., Nickerson, D.A., Anderson, A.D., Weir, B.S., Livingston, R.J., Dean, M.D., Smith, K.L., Schadt, E.E., Nachman, M.W., 2007. Linkage disequilibrium in wild mice. PLoS Genet. 3, 1487–1495. http://dx.doi.org/10.1371/journal.pgen.0030144.

Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Yang, J.L., Ferrante, T.C., Terry, R., Jeanty, S.S.F., Li, C., Amamoto, R., Peters, D.T., Turczyk, B.M., Marblestone, A.H., Inverso, S.A., Bernard, A., Mali, P., Rios, X., Aach, J., Church, G.M., 2014. Highly multiplexed subcellular RNA sequencing in situ. Science 343, 1360–1363. http://dx.doi.org/10.1126/science.1250212.

Lee, J.-H., Park, I.-H., Gao, Y., Li, J.B., Li, Z., Daley, G.Q., Zhang, K., Church, G.M., 2009. A robust approach to identifying tissue-specific gene expression regulatory variants using personalized human induced pluripotent stem cells. PLoS Genet. 5, e1000718. http://dx.doi.org/10.1371/journal.pgen.1000718.

Lei, K.F., 2012. Microfluidic systems for diagnostic applications: a review. J. Lab. Autom. 17, 330–347. http://dx.doi.org/10.1177/2211068212454853.

Li, H.H., Gyllensten, U.B., Cui, X.F., Saiki, R.K., Erlich, H.A., Arnheim, N., 1988. Amplification and analysis of DNA sequences in

single human sperm and diploid cells. Nature 335, 414–417. http://dx.doi.org/10.1038/335414a0.

Li, J., Young, C.S.H., 2005. In situ detection of specific DNA double strand breaks using rolling circle amplification. Cell Cycle 4, 1767–1773. http://dx.doi.org/10.4161/cc.4.12.2211.

Li, J.B., Levanon, E.Y., Yoon, J.-K., Aach, J., Xie, B., Leproust, E., Zhang, K., Gao, Y., Church, G.M., 2009a. Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. Science 324, 1210–1213. http://dx.doi.org/10.1126/science.1170995.

Li, R., Li, Y., Fang, X., Yang, H., Wang, J., Kristiansen, K., Wang, J., 2009b. SNP detection for massively parallel whole-genome resequencing. Genome Res. 19, 1124–1132. http://dx.doi.org/10.1101/gr.088013.108.

Liu, D., Daubendiek, S.L., Zillman, M.A., Ryan, K., Kool, E.T., 1996. Rolling circle DNA synthesis: small circular oligonucleotides as efficient templates for DNA polymerases. J. Am. Chem. Soc. 118, 1587–1594. http://dx.doi.org/10.1021/ja952786k.

Lizardi, P.M., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D.C., Ward, D.C., 1998. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. Nat. Genet. 19, 225–232. http://dx.doi.org/10.1038/898.

Lohmann, J.S., Stougaard, M., Koch, J., 2007. Detection of short repeated genomic sequences on metaphase chromosomes using padlock probes and target primed rolling circle DNA synthesis. BMC Mol. Biol. 8, 103. http://dx.doi.org/10.1186/1471-2199-8-103.

Lyamichev, V., Brow, M.A., Dahlberg, J.E., 1993. Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. Science 260, 778–783. http://dx.doi.org/10.1126/science.7683443.

Mahmoudian, L., Kaji, N., Tokeshi, M., Nilsson, M., Baba, Y., 2008. Rolling circle amplification and circle-to-circle amplification of a specific gene integrated with electrophoretic analysis on a single chip. Anal. Chem. 80, 2483–2490. http://dx.doi.org/10.1021/ac702289j.

Mardis, E.R., 2008. The impact of next-generation sequencing technology on genetics. Trends Genet. 24, 133–141. http://dx.doi.org/10.1016/j.tig.2007.12.007.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L.I., Jarvie, T.P., Jirage, K.B., Kim, J.-B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., Rothberg, J.M., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376–380. http://dx.doi.org/10.1038/nature04726.

Maruyama, F., Kenzaka, T., Yamaguchi, N., Tani, K., Nasu, M., 2005. Visualization and enumeration of bacteria carrying a specific gene sequence by in situ rolling circle amplification. Appl. Environ. Microbiol. 71, 7933–7940. http://dx.doi.org/10.1128/AEM.71.12.7933-7940.2005.

Maruyama, F., Tani, K., Kenzaka, T., Yamaguchi, N., Nasu, M., 2006. Quantitative determination of free-DNA uptake in river bacteria at the single-cell level by in situ rolling-circle amplification. Appl. Environ. Microbiol. 72, 6248–6256. http://dx.doi.org/10.1128/AEM.03035-05.

McCarthy, E.L., Bickerstaff, L.E., Pereira da Cunha, M., Millard, P.J., 2007. Nucleic acid sensing by regenerable surface-associated isothermal rolling circle amplification. Biosens. Bioelectron. 22, 1236–1244. http://dx.doi.org/10.1016/j.bios.2006.05.001.

McCarthy, E.L., Egeler, T.J., Bickerstaff, L.E., Pereira da Cunha, M., Millard, P.J., 2006. Detection and identification of IHN and ISA viruses by isothermal DNA amplification in microcapillary tubes. Anal. Bioanal. Chem. 386, 1975–1984. http://dx.doi.org/10.1007/s00216-006-0872-9.

Melin, J., Jarvius, J., Göransson, J., Nilsson, M., 2007. Homogeneous amplified single-molecule detection: characterization of key parameters. Anal. Biochem. 368, 230–238. http://dx.doi.org/10.1016/j.ab.2007.05.001.

Mezger, A., Gullberg, E., Göransson, J., Zorzet, A., Herthnek, D., Tano, E., Nilsson, M., Andersson, D.I., 2015a. A general method for rapid determination of antibiotic susceptibility and species in bacterial infections. J. Clin. Microbiol. 53, 425–432. http://dx.doi.org/10.1128/JCM.02434-14.

Mezger, A., Kühnemund, M., Nilsson, M., Herthnek, D., 2015b. Highly specific DNA detection employing ligation on suspension bead array readout. N. Biotechnol. 32, 504–510. http://dx.doi.org/10.1016/j.nbt.2015.01.011.

Mezger, A., Öhrmalm, C., Herthnek, D., Blomberg, J., Nilsson, M., 2014. Detection of rotavirus using padlock probes and rolling circle amplification. PLoS One 9, e111874. http://dx.doi.org/10.1371/journal.pone.0111874.

Millard, P.J., Bickerstaff, L.E., LaPatra, S.E., Kim, C.H., 2006. Detection of infectious haematopoietic necrosis virus and infectious salmon anaemia virus by molecular padlock amplification. J. Fish. Dis. 29, 201–213. http://dx.doi.org/10.1111/j.1365-2761.2006.00705.x.

Moens, L.N., Falk-Sörqvist, E., Asplund, A.C., Bernatowska, E., Smith, C.I.E., Nilsson, M., 2014. Diagnostics of primary immunodeficiency diseases: a sequencing capture approach. PLoS One 9, e114901. http://dx.doi.org/10.1371/journal.pone.0114901.

Moens, L.N.J., Falk-Sörqvist, E., Ljungström, V., Mattsson, J., Sundström, M., La Fleur, L., Mathot, L., Micke, P., Nilsson, M., Botling, J., 2015. HaloPlex targeted resequencing for mutation detection in clinical formalin-fixed, paraffin-embedded tumor samples. J. Mol. Diagn. http://dx.doi.org/10.1016/j.jmoldx.2015.06.009.

Najafzadeh, M.J., Sun, J., Vicente, V.A., De Hoog, G.S., 2011. Rapid identification of fungal pathogens by rolling circle amplification using Fonsecaea as a model. Mycoses 54, e577–e582. http://dx.doi.org/10.1111/j.1439-0507.2010.01995.x.

Nilsson, M., Antson, D.O., Barbany, G., Landegren, U., 2001. RNA-templated DNA ligation for transcript analysis. Nucleic Acids Res. 29, 578–581. http://dx.doi.org/10.1093/nar/29.2.578.

Nilsson, M., Barbany, G., Antson, D.O., Gertow, K., Landegren, U., 2000. Enhanced detection and distinction of RNA by enzymatic probe ligation. Nat. Biotechnol. 18, 791–793. http://dx.doi.org/10.1038/77367.

Nilsson, M., Krejci, K., Koch, J., Kwiatkowski, M., Gustavsson, P., Landegren, U., 1997. 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. http://dx.doi.org/10.1038/ng0797-252.

Nilsson, M., Malmgren, H., Samiotaki, M., Kwiatkowski, M., Chowdhary, B.P., Landegren, U., 1994. Padlock probes: circularizing oligonucleotides for localized DNA detection. Science 265, 2085–2088. http://dx.doi.org/10.1126/science.7522346.

Novais, R.C., Borsuk, S., Dellagostin, O.A., Thorstenson, Y.R., 2008. Molecular inversion probes for sensitive detection of *Mycobacterium tuberculosis*. J. Microbiol. Methods 72, 60–66. http://dx.doi.org/10.1016/j.mimet.2007.10.009.

Okou, D.T., Steinberg, K.M., Middle, C., Cutler, D.J., Albert, T.J., Zwick, M.E., 2007. Microarray-based genomic selection for high-throughput resequencing. Nat. Methods 4, 907–909. http://dx.doi.org/10.1038/nmeth1109.

Østerberg, F.W., Rizzi, G., Donolato, M., Bejhed, R.S., Mezger, A., Strömberg, M., Nilsson, M., Strømme, M., Svedlindh, P., Hansen, M.F., 2014. On-chip detection of rolling circle amplified DNA molecules from *Bacillus globigii* spores and *Vibrio cholerae*. Small 10, 2877–2882. http://dx.doi.org/10.1002/smll.201303325.

Østerberg, F.W., Rizzi, G., Zardán Gómez de la Torre, T., Strömberg, M., Strømme, M., Svedlindh, P., Hansen, M.F., 2013. Measurements of Brownian relaxation of magnetic nanobeads using planar Hall effect bridge sensors. Biosens. Bioelectron. 40, 147–152. http://dx.doi.org/10.1016/j.bios.2012.07.002.

Pickering, J., Bamford, A., Godbole, V., Briggs, J., Scozzafava, G., Roe, P., Wheeler, C., Ghouze, F., Cuss, S., 2002. Integration of DNA ligation and rolling circle amplification for the homogeneous, end-point detection of single nucleotide polymorphisms. Nucleic Acids Res. 30, e60. http://dx.doi.org/10.1093/nar/gnf060.

Porreca, G.J., Shendure, J., Church, G.M., 2006. Polony DNA sequencing. Curr. Protoc. Mol. Biol. http://dx.doi.org/10.1002/0471142727.mb0708s76 (Chapter 7, Unit 7.8).

Porreca, G.J., Zhang, K., Li, J.B., Xie, B., Austin, D., Vassallo, S.L., LeProust, E.M., Peck, B.J., Emig, C.J., Dahl, F., Gao, Y., Church, G.M., Shendure, J., 2007. Multiplex amplification of large sets of human exons. Nat. Methods 4, 931–936. http://dx.doi.org/10.1038/nmeth1110.

Roberts, G.S., Kozak, D., Anderson, W., Broom, M.F., Vogel, R., Trau, M., 2010. Tunable nano/micropores for particle detection and discrimination: scanning ion occlusion spectroscopy. Small 6, 2653–2658. http://dx.doi.org/10.1002/smll.201001129.

Russell, C., Welch, K., Jarvius, J., Cai, Y., Brucas, R., Nikolajeff, F., Svedlindh, P., Nilsson, M., 2014. Gold nanowire based electrical DNA detection using rolling circle amplification. ACS Nano 8, 1147–1153. http://dx.doi.org/10.1021/nn4058825.

Saiyed, Z., Telang, S., Ramchand, C., 2003. Application of magnetic techniques in the field of drug discovery and biomedicine. Biomagn. Res. Technol. 1, 2. http://dx.doi.org/10.1186/1477-044X-1-2.

Salas, M., Vega, M., de, Lázaro, J.M., Blanco, L., 2004. Phi29 DNA Polymerase, a Potent Amplification Enzyme.

Salmon Hillbertz, N.H.C., Isaksson, M., Karlsson, E.K., Hellmén, E., Pielberg, G.R., Savolainen, P., Wade, C.M., von Euler, H., Gustafson, U., Hedhammar, A., Nilsson, M., Lindblad-Toh, K., Andersson, L., Andersson, G., 2007. Duplication of FGF3, FGF4, FGF19 and ORAOV1 causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs. Nat. Genet. 39, 1318–1320. http://dx.doi.org/10.1038/ng.2007.4.

Sato, K., Ishii, R., Sasaki, N., Sato, K., Nilsson, M., 2013. Bead-based padlock rolling circle amplification for single DNA molecule counting. Anal. Biochem. 437, 43–45. http://dx.doi.org/10.1016/j.ab.2013.02.016.

Sato, K., Tachihara, A., Renberg, B., Mawatari, K., Sato, K., Tanaka, Y., Jarvius, J., Nilsson, M., Kitamori, T., 2010. Microbead-based rolling circle amplification in a microchip for sensitive DNA detection. Lab. Chip 10, 1262–1266. http://dx.doi.org/10.1039/b927460j.

Schopf, E., Liu, Y., Deng, J.C., Yang, S., Cheng, G., Chen, Y., 2011. *Mycobacterium tuberculosis* detection via rolling circle amplification. Anal. Methods 3, 267. http://dx.doi.org/10.1039/c0ay00529k.

Shaposhnikov, S., Larsson, C., Henriksson, S., Collins, A., Nilsson, M., 2006. Detection of Alu sequences and mtDNA in comets using padlock probes. Mutagenesis 21, 243–247. http://dx.doi.org/10.1093/mutage/gel022.

Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D., Church, G.M., 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. Science 309, 1728–1732. http://dx.doi.org/10.1126/science.1117389.

Smolina, I.V., Kuhn, H., Lee, C., Frank-Kamenetskii, M.D., 2008. Fluorescence-based detection of short DNA sequences under non-denaturing conditions. Bioorg. Med. Chem. 16, 84–93. http://dx.doi.org/10.1016/j.bmc.2007.04.063.

Smolina, I., Lee, C., Frank-Kamenetskii, M., 2007. Detection of low-copy-number genomic DNA sequences in individual bacterial cells by using peptide nucleic acid-assisted rolling-circle amplification and fluorescence in situ hybridization. Appl. Environ. Microbiol. 73, 2324–2328. http://dx.doi.org/10.1128/AEM.02038-06.

Smyth, D.J., Cooper, J.D., Bailey, R., Field, S., Burren, O., Smink, L.J., Guja, C., Ionescu-Tirgoviste, C., Widmer, B., Dunger, D.B., Savage, D.A., Walker, N.M., Clayton, D.G., Todd, J.A., 2006. A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. Nat. Genet. 38, 617–619. http://dx.doi.org/10.1038/ng1800.

Steain, M.C., Dwyer, D.E., Hurt, A.C., Kol, C., Saksena, N.K., Cunningham, A.L., Wang, B., 2009. Detection of influenza A H1N1 and H3N2 mutations conferring resistance to oseltamivir using rolling circle amplification. Antiviral Res. 84, 242–248. http://dx.doi.org/10.1016/j.antiviral.2009.09.010.

Stoddard, J.L., Niemela, J.E., Fleisher, T.A., Rosenzweig, S.D., 2014. Targeted NGS: a cost-effective approach to molecular diagnosis of PIDs. Front. Immunol. 5 http://dx.doi.org/10.3389/fimmu.2014.00531.

Stougaard, M., Lohmann, J.S., Zajac, M., Hamilton-Dutoit, S., Koch, J., 2007. In situ detection of non-polyadenylated RNA molecules using Turtle Probes and target primed rolling circle PRINS. BMC Biotechnol. 7, 69. http://dx.doi.org/10.1186/1472-6750-7-69.

Strömberg, M., Göransson, J., Gunnarsson, K., Nilsson, M., Svedlindh, P., Strømme, M., 2008. Sensitive molecular diagnostics using volume-amplified magnetic nanobeads. Nano Lett. 8, 816–821. http://dx.doi.org/10.1021/nl072760e.

Strömberg, M., Zardán Gómez de la Torre, T., Göransson, J., Gunnarsson, K., Nilsson, M., Svedlindh, P., Strømme, M., 2009. Multiplex detection of DNA sequences using the volume-amplified magnetic nanobead detection assay. Anal. Chem. 81, 3398–3406. http://dx.doi.org/10.1021/ac900561r.

Strömberg, M., Zardán Gómez de la Torre, T., Nilsson, M., Svedlindh, P., Strømme, M., 2014. A magnetic nanobead-based bioassay provides sensitive detection of single- and biplex bacterial DNA using a portable AC susceptometer. Biotechnol. J. 9, 137–145. http://dx.doi.org/10.1002/biot.201300348.

Tanaka, Y., Xi, H., Sato, K., Mawatari, K., Renberg, B., Nilsson, M., Kitamori, T., 2011. Single-molecule DNA patterning and detection by padlock probing and rolling circle amplification in microchannels for analysis of small sample volumes. Anal. Chem. 83, 3352–3357. http://dx.doi.org/10.1021/ac103185j.

Thomas, D.C., Nardone, G.A., Randall, S.K., 1999. Amplification of padlock probes for DNA diagnostics by cascade rolling circle amplification or the polymerase chain reaction. Arch. Pathol. Lab. Med. 123, 1170–1176. http://dx.doi.org/10.1043/1543-2165-123.20.1170.

Thorsen, T., Maerkl, S.J., Quake, S.R., 2002. Microfluidic large-scale integration. Science 298, 580–584. http://dx.doi.org/10.1126/science.1076996.

Tian, J., Gong, H., Sheng, N., Zhou, X., Gulari, E., Gao, X., Church, G., 2004. Accurate multiplex gene synthesis from programmable DNA microchips. Nature 432, 1050–1054. http://dx.doi.org/10.1038/nature03151.

Tong, Z., Kong, F., Wang, B., Zeng, X., Gilbert, G.L., 2007. A practical method for subtyping of Streptococcus agalactiae serotype III, of human origin, using rolling circle amplification. J. Microbiol. Methods 70, 39–44. http://dx.doi.org/10.1016/j.mimet.2007.03.010.

Umbarger, M.A., Kennedy, C.J., Saunders, P., Breton, B., Chennagiri, N., Emhoff, J., Greger, V., Hallam, S., Maganzini, D., Micale, C., Nizzari, M.M., Towne, C.F., Church, G.M., Porreca, G.J., 2014. Next-generation carrier screening. Genet. Med. 16, 132–140.

van Doorn, R., Sławiak, M., Szemes, M., Dullemans, A.M., Bonants, P., Kowalchuk, G.A., Schoen, C.D., 2009. Robust detection and identification of multiple oomycetes and fungi in environmental samples by using a novel cleavable padlock probe-based ligation detection assay. Appl. Environ. Microbiol. 75, 4185–4193. http://dx.doi.org/10.1128/AEM.00071-09.

van Doorn, R., Szemes, M., Bonants, P., Kowalchuk, G.A., Salles, J.F., Ortenberg, E., Schoen, C.D., 2007. Quantitative multiplex detection of plant pathogens using a novel ligation probe-based system coupled with universal, high-throughput real-time PCR on OpenArrays. BMC Genomics 8, 276. http://dx.doi.org/10.1186/1471-2164-8-276.

van Eijk, M.J.T., Broekhof, J.L.N., van der Poel, H.J.A., Hogers, R.C.J., Schneiders, H., Kamerbeek, J., Verstege, E., van Aart, J.W., Geerlings, H., Buntjer, J.B., van Oeveren, A.J., Vos, P., 2004. SNPWave: a flexible multiplexed SNP genotyping technology. Nucleic Acids Res. 32, e47. http://dx.doi.org/10.1093/nar/gnh045.

Vogel, R., Willmott, G., Kozak, D., Roberts, G.S., Anderson, W., Groenewegen, L., Glossop, B., Barnett, A., Turner, A., Trau, M., 2011. Quantitative sizing of nano/microparticles with a tunable elastomeric pore sensor. Anal. Chem. 83, 3499–3506. http://dx.doi.org/10.1021/ac200195n.

Vogelstein, B., Kinzler, K.W., 1999. Digital PCR. Proc. Natl. Acad. Sci. U.S.A. 96, 9236–9241. http://dx.doi.org/10.1073/pnas.96.16.9236.

Wamsley, H.L., Barbet, A.F., 2008. In situ detection of *Anaplasma* spp. by DNA target-primed rolling-circle amplification of a padlock probe and intracellular colocalization with immunofluorescently labeled host cell von Willebrand factor. J. Clin. Microbiol. 46, 2314–2319. http://dx.doi.org/10.1128/JCM.02197-07.

Wang, B., Dwyer, D.E., Blyth, C.C., Soedjono, M., Shi, H., Kesson, A., Ratnamohan, M., McPhie, K., Cunningham, A.L., Saksena, N.K., 2010. Detection of the rapid emergence of the H275Y mutation associated with oseltamivir resistance in severe pandemic influenza virus A/H1N1 09 infections. Antiviral Res. 87, 16–21. http://dx.doi.org/10.1016/j.antiviral.2010.04.002.

Wang, F., Flanagan, J., Su, N., Wang, L.-C., Bui, S., Nielson, A., Wu, X., Vo, H.-T., Ma, X.-J., Luo, Y., 2012. RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. J. Mol. Diagn. 14, 22–29. http://dx.doi.org/10.1016/j.jmoldx.2011.08.002.

Wang, Y., Moorhead, M., Karlin-Neumann, G., Falkowski, M., Chen, C., Siddiqui, F., Davis, R.W., Willis, T.D., Faham, M., 2005. Allele quantification using molecular inversion probes (MIP). Nucleic Acids Res. 33, e183. http://dx.doi.org/10.1093/nar/gni177.

Zhang, D.Y., Brandwein, M., Chun Hung Hsuih, T., Li, H., 1998. Amplification of target-specific, ligation-dependent circular probe. Gene 211, 277–285. http://dx.doi.org/10.1016/S0378-1119(98)00113-9.

Zanke, B.W., Greenwood, C.M.T., Rangrej, J., Kustra, R., Tenesa, A., Farrington, S.M., Prendergast, J., Olschwang, S., Chiang, T., Crowdy, E., Ferretti, V., Laflamme, P., Sundararajan, S., Roumy, S., Olivier, J.-F., Robidoux, F., Sladek, R., Montpetit, A., Campbell, P., Bezieau, S., O'Shea, A.M., Zogopoulos, G., Cotterchio, M., Newcomb, P., McLaughlin, J., Younghusband, B., Green, R., Green, J., Porteous, M.E.M., Campbell, H., Blanche, H., Sahbatou, M., Tubacher, E., Bonaiti-Pellié, C., Buecher, B., Riboli, E., Kury, S., Chanock, S.J., Potter, J., Thomas, G., Gallinger, S., Hudson, T.J., Dunlop, M.G., 2007. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat. Genet. 39, 989–994. http://dx.doi.org/10.1038/ng2089.

Zhang, K., Li, J.B., Gao, Y., Egli, D., Xie, B., Deng, J., Li, Z., Lee, J.-H., Aach, J., Leproust, E.M., Eggan, K., Church, G.M., 2009. Digital RNA allelotyping reveals tissue-specific and allele-specific gene expression in human. Nat. Methods 6, 613–618. http://dx.doi.org/10.1038/nmeth.1357.

Zheng, S.L., Liu, W., Wiklund, F., Dimitrov, L., Bälter, K., Sun, J., Adami, H.O., Johansson, J.E., Sun, J., Chang, B., Loza, M., Turner, A.R., Bleecker, E.R., Meyers, D.A., Carpten, J.D., Duggan, D., Isaacs, W.B., Xu, J., Grönberg, H., 2006. A comprehensive association study for genes in inflammation pathway provides support for their roles in prostate cancer risk in the CAPS study. Prostate 66, 1556–1564. http://dx.doi.org/10.1002/pros.20496.

Zhong, X.B., Lizardi, P.M., Huang, X.H., Bray-Ward, P.L., Ward, D.C., 2001. Visualization of oligonucleotide probes and point mutations in interphase nuclei and DNA fibers using rolling circle DNA amplification. Proc. Natl. Acad. Sci. U.S.A. 98, 3940–3945. http://dx.doi.org/10.1073/pnas.061026198.

Chapter 11

Advances in Microfluidics and Lab-on-a-Chip Technologies

H. Jayamohan, V. Romanov, H. Li, J. Son, R. Samuel, J. Nelson and B.K. Gale
University of Utah, Salt Lake City, UT, United States

11.1 OVERVIEW

Microfluidics has come a long way since the seminal paper by Andreas Manz in 1990 (Manz, 1990). The paper envisioned an integrated, automated platform for performing a range of analysis steps. Advances since then have brought us closer to the vision of a sample-in, answer-out platform (Jenkins and Mansfield, 2013; Lee, 2013). At the turn of the century, the introduction of polydimethylsiloxane (PDMS) and soft lithography gave a major boost to the field (Duffy et al., 1998; Unger et al., 2000). Soft lithography and other inexpensive and rapid microfabrication techniques have enabled researchers to apply microfluidics to a wide range of areas such as catalysis, molecular point-of-care (POC) diagnostics remains the primary application area of microfluidics (Jayamohan et al., 2013) and it accounts for the largest share of the microfluidics market (Yetisen and Volpatti, 2014).

Microfluidics can be broadly defined as systems leveraging micrometer scale channels to manipulate and process low-volume (10^{-9} to 10^{-18} L) fluid samples (Whitesides, 2006). Such systems enable advantages such as the capability of processing low volumes of samples requiring lower amounts of expensive reagents. Lab-on-a-chip (LOC) platforms leveraging microfluidics are capable of carrying out separations and detections with high resolution and sensitivity. The smaller length scales associated with microfluidics enable faster analysis and reduced response times. Advances in microfluidic manufacturing methods (lithography, xurography, and laser machining) (Jayamohan et al., 2013) have enabled devices with a smaller footprint at a reduced cost. This is especially important for the POC applicability of microfluidic devices in the context of global health. The convergence of microfluidics with nanotechnology-based barcode techniques (quantum dots) (Klostranec et al., 2007) and oligonucleotide labels (Jayamohan et al., 2015) (metal nanoparticles) has enabled the multiplexed ultrasensitive detection of analytes from complex sample matrices involving contaminants (Hauck et al., 2010; Derveaux et al., 2008; Sanvicens et al., 2009).

Many microfluidic platforms are limited in their application and adoption by requirements involving the need for significant off-chip sample preparation. Developments in on-chip sample preparation have offset some of these challenges. Microfluidic systems also have challenges owing to scaling such as capillary forces, surface roughness, air bubbles (Lochovsky et al., 2012), surface fouling (Schoenitz et al., 2015), channel clogging, and laminar flow-limiting reagent mixing to diffusion. Other issues relate to volume mismatch between real-world samples and microfluidic components, and interfacing of electronics and fluids on the microscale (Fredrickson and Fan, 2004).

Despite the significant academic interest in microfluidics, commercial applications have not evolved at a similar rate (Chin et al., 2012). The success of materials such as PDMS in microfluidics academic research has not translated well to industry owing to issues with manufacturability and scaling. Also, there is a lack of academic research on microfluidic devices fabricated using alternative materials (glass and thermoplastic polymer), which has prevented the rapid transfer of these technologies from the laboratory to the market (Yetisen and Volpatti, 2014). Microfluidic commercialization is also limited because of the customized nature of each assay or microfluidic chip: there is no universal fabrication approach that can be implemented in most needed applications. Another area of concern is the lack of statistical reproducibility and microfluidic chip-to-chip variability among published research (Becker, 2010).

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00011-0
Copyright © 2017 Elsevier Ltd. All rights reserved.

New processes such as droplet (emulsion) and paper microfluidics seem to be overcoming some of these challenges with increasing adoption by both industry and researchers alike (Lee, 2013; Hindson et al., 2011). Droplet microfluidics use two immiscible fluids to establish compartmentalization within pico- or nanoliter-sized droplets (Teh et al., 2008). Paper microfluidics replaces hollow, free-flow microchannels with woven microfibers of paper that wick fluids, circumventing the need for additional pumps (Lee, 2013) but giving up some flexibility. As might be imagined, microfluidics printed on paper can be relatively inexpensive. Looking forward, three-dimensional (3D) printing holds promise in extending these capabilities to other materials, including plastics, for microfluidic device development.

Overall, microfluidic approaches to a wide variety of molecular diagnostics applications are developing rapidly. In this chapter, we will briefly review some of the most important and most impactful applications of microfluidics in molecular diagnostics. Applications in nucleic acids, proteins, cell preparation for molecular diagnostics, and other targets will be discussed briefly.

11.2 MICROFLUIDICS FOR DNA AMPLIFICATION AND ANALYSIS

DNA analysis and amplification are becoming standard practice in many diagnostic and analytical procedures, with polymerase chain reaction (PCR) being one of the most robust and popular molecular diagnostic techniques in medicine (Chang et al., 2013). DNA amplification techniques can be broadly categorized as isothermal and nonisothermal. Isothermal DNA amplification techniques are carried out at a constant temperature and tend to be simpler mechanically, so interest in this area is high, leading to a number of isothermal DNA amplification techniques being reported in the past couple of years (Chang et al., 2013) (see also Chapter 3).

Isothermal DNA amplification techniques are well-suited for microfluidic integration in applications where reasonably fast (15- to 60-min) DNA amplification is needed in low-resource settings, as temperature cycling is not needed, which significantly simplifies the hardware needed to carry out isothermal DNA amplification. Table 11.1 summarizes some promising isothermal DNA

TABLE 11.1 Promising Isothermal DNA Amplification Techniques for Incorporation in Microfluidic Systems

Reaction	Type of Template Required	Reaction Temperature (°C)	Highlights/Comments	Multiplex Capability
Nucleic acid sequence-based amplification	RNA	41	**1.** Prone to nonspecific amplification **2.** Requires initial heating of template RNA at 65°C	Yes
Loop-mediated isothermal amplification (LAMP)	ss-DNA	60–65	**1.** Using two primer sets, the LAMP reaction becomes specific **2.** Requires careful design of primer sets **3.** Ease of detection of amplified products owing to production of pyrophosphate (visible to naked eye) as by-product of positive LAMP reaction	Yes
Helicase-dependent amplification (HDA)	ds-DNA	45–65	**1.** Uses single primer set, which makes HDA simple process with ease of optimization. However, the speed of HDA is low when samples contain <100 DNA copies, but optimizing the reaction for a specific amplicon can compensate for this	Two-plex
Strand-displacement amplification	ss-DNA	37–70	**1.** Requires initial heating of template DNA at 95°C **2.** Prone to nonspecific amplification. Slow reaction	Yes
Recombinase polymerase amplification	ds-DNA	37–42	**1.** Fast reaction (probably one of the fastest among other isothermal DNA amplification techniques). Robust reaction without requiring precise temperature control	Yes

ds-DNA, double-stranded DNA; *ss-DNA*, single-stranded DNA.

amplification techniques that have been successfully demonstrated in microfluidic systems in commercial and academic settings. For further details readers can refer to reviews by Asiello and co-workers and others (Asiello and Baeumner, 2011; Craw and Balachandran, 2012; Tröger et al., 2015).

PCR is the predominant and most popular non-isothermal DNA amplification technique and is used in many microfluidic devices. PCR involves three substeps that occur at different temperatures. PCR typically requires at least 35–40 temperature cycles for a single PCR reaction to achieve useful concentrations. The speed at which PCR can be run depends on two factors: the speed of the DNA polymerase and the heat transfer rate of the hardware performing the PCR. Because microfluidic systems are inherently small, leading to a small thermal mass, and having a high surface-to-volume ratio, they are naturally capable of providing rapid heat transfer rates. Microfluidic PCR systems also offer the ability to automate the preparation of the PCR reaction mix, thereby reducing the risk of contamination and false positives by human error. Finally, microfluidic PCR systems require low sample volumes, which are helpful when the genetic material being tested is scarce, and the reagent volumes are likewise low, significantly reducing costs.

Since the inception of microfluidics in the 1980s, a considerable amount of work has been done to develop microfluidic devices for PCR. There are generally two types of microfluidic systems for PCR: flow-through, stationary, and droplet digital (Chang et al., 2013). In a flow-through PCR system, the PCR mixture travels through a microchannel that contains temperature regions for all three substeps of PCR. In some versions, the sample may be moved back and forth between the temperature regions whereas in others the sample reaches the temperatures by continually moving forward. In stationary PCR systems the PCR mixture remains stationary in a microchamber while the temperature of the microchamber cycles through the needed temperatures. There are many variations of these approaches. For example, in droplet digital PCR systems, the PCR reaction mix along with the template DNA is encapsulated in a microdroplet and then is transported to different regions of a microchip or is temperature cycled in place (Prakash et al., 2014).

Thousands of microfluidic PCR devices have been demonstrated successfully with measurable real-time amplification incorporated in the microfluidic PCR chip; some show amplification completed in a few minutes (Chang et al., 2013; Pješčić et al., 2010; Crews et al., 2008; Neuzil et al., 2006) even at the single-cell level (Zhu et al., 2012). For example, Fig. 11.1 shows a microfluidic chip performing both continuous PCR and high-resolution melting analysis (HRMA) simultaneously in less than 6 min for 30 cycles. The PCR is progressing down the image while HRMA can be performed simultaneously for each cycle in the horizontal direction by measuring the fluorescence intensity in the image. A similar chip has been shown to complete PCR in less than 1 min (Samuel et al., 2016) using extreme PCR (Farrar and Wittwer, 2015). Furthermore, biomedical diagnostic companies have commercialized several microfluidic PCR systems (Cao et al., 2015; Volpatti and Yetisen, 2014).

11.2.1 DNA Sequencing and Mutation Detection

In cancer and other diseases, altered DNA gene patterns or mutations have been found to be useful biomarkers for the detection and diagnosis of disease (Almoguera et al., 1988). Detecting mutations requires the ability to sequence at least a small part of a genetic sequence, which has led to major efforts to develop high-speed, high-throughput DNA sequencing methods. As microfluidics has emerged as a tool for clinical molecular diagnostics, applications in

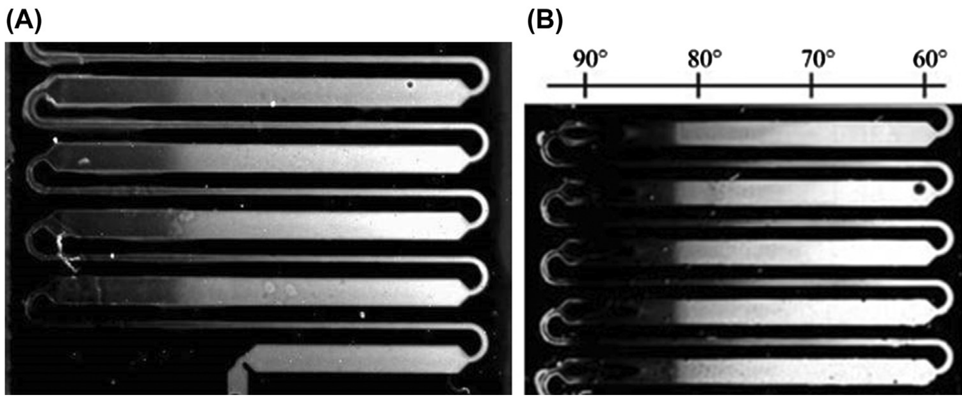


FIGURE 11.1 Spatial continuous-flow polymerase chain reaction (PCR) showing both PCR and high-resolution melting analysis melting. (A) The image shows a microfluidic chip design that has two types of channel widths. Denaturation of DNA occurs in narrow channels, whereas annealing and extension (which are relatively slower than denaturation) occur in the wider channels owing to the reduced flow rate. (B) The temperature zones along the chip are labeled (°C).

mutation detection and genetic screening have developed with the promise of profiling genetic sequences quickly and interpreting the implications of such sequences. Traditional macro-DNA sequencing includes steps such as cell preparation, amplification, purification, and electrophoresis. Each step can be scaled down and integrated into a microfluidic device to achieve rapid and low-cost DNA sequencing (Paegel et al., 2003). Other nontraditional approaches, often adapting macroscale methods, for detecting altered gene sequences or sequencing short sections of genes have been developed, including digital PCR and HRMA. The application of microfluidic technology to many of these sequencing or mutation detection techniques is discussed subsequently.

11.2.2 Capillary Electrophoresis

In 1995, Woolley and Mathies developed a microfabricated capillary electrophoresis (CE) chip that can complete DNA sequencing with 97% accuracy and approximately 150 bases in 540 s for four-color separations. The CE chip demonstrated the feasibility of fast and high-throughput DNA sequencing (Woolley and Mathies, 1995). In 1999, Liu et al. presented an improved microfabricated CE chip. The separation matrix, temperature, channel and injector size, and injector parameters were all optimized to achieve better DNA sequencing performance. The optimized chip could achieve about 500 bases in 20 min for four-color separations (Liu et al., 1999). Paegel et al. (2002) developed a radially symmetric, 96-lane capillary array electrophoresis chip that acquired approximately 41,000 bases in only 24 min. Similar approaches and further improvement studies were summarized in Paegel et al. (2003) and these approaches are regularly used. This topic was reviewed in depth in a previous version of this book (Jayamohan et al., 2013). As an example of what has been accomplished, micro-CE integrated systems have been used for quantitative detection of low-abundance mutations of the KRAS gene from paraffin tissue sections of colorectal cancer. These systems have nanoliter sample introduction components leading to CE separation of the target genes and detection by laser-induced fluorescence, all of which is accomplished in minutes or even seconds (Zhang et al., 2013; Xu et al., 2010).

11.2.3 DNA Purification

He et al. (2000) used capillary zone electrophoresis to purify sequencing fragments. Khandurina et al. (2002) developed a microfluidic device for fraction collection of various size DNA fragments. In Tian's study, the effectiveness of a variety of silica resin for miniaturized DNA purification was evaluated (Tian et al., 2000). Other alternative on-chip approaches for DNA purification were studied, including using temperature gradients along the channel and hybridization-mediated capture (Paegel et al., 2003). More recent techniques not specifically for sequencing but generally relevant are discussed in the sample preparation section.

11.2.4 DNA Amplification and Sanger Sequencing

Many studies performing low-volume Sanger cycle sequencing in microfluidics have been presented. Hadd et al. (2000) and Xue et al. (2001) demonstrated low-volume reactions inside a capillary, which established the feasibility of small-scale sample preparation. Lagally et al. (2000) developed the first nanoliter-scale DNA amplification system, which was integrated with electrophoretic analysis on a microfluidic chip. In 2006, Blazej et al. developed a microfabricated bioprocessor to integrate all three Sanger sequencing steps. This microdevice was built on a hybrid glass-PDMS wafer and enables complete Sanger sequencing from 1 fM of DNA template within 30 min. With further improvements, the starting template for DNA sequencing has been reduced from 1 fM to 100 aM (Blazej et al., 2006). In 2013, Abate et al. developed a droplet-based microfluidic system for DNA sequencing in a rapid and inexpensive manner.

11.2.5 Digital Polymerase Chain Reaction

Although PCR has long been used for mutation detection, a highly sensitive version of PCR, digital PCR, has been gaining a significant following both commercially and scientifically. Droplet-based digital PCR puts the PCR milieu into thousands or millions of drops with the general goal of keeping any amplification targets at less than one per drop, which results in an "on" or "off" signal for each drop when the PCR is complete. Digital PCR can reduce overall analysis costs and reduces sample cross-contamination. Most important, droplet-based detection can provide a highly sensitive and high-throughput method for detecting DNA mutations (Hsieh et al., 2009; Pekin et al., 2011). As an example, droplet-based digital PCR enabled the precise determination of mutations in several cancer cell lines and the precise quantification of a single mutated KRAS gene in a background of 200,000 unmutated KRAS genes (Pekin et al., 2011). In a similar vein, BRAF mutation detection was accomplished using a spinning disk digital PCR layout (Fig. 11.2) (Sundberg et al., 2014). In this case a polymer microfluidic system the size of a DVD was fabricated with a single spiral channel that contained 1000 microwells on the outside of the disc. After injecting a PCR reaction mix, the disc is spun and a combination of centrifugal forces and the laminar flow profile distribute the reaction mixture into the wells. An oil plug is

FIGURE 11.2 Spinning disk for digital polymerase chain reaction (PCR). (*Inset*) Post-PCR fluorescent image of the disk. Fluorescence correlates with the number of copies per well. *Reprinted with permission from Sundberg, S.O., et al., 2014. Quasi-digital PCR: enrichment and quantification of rare DNA variants. Biomed. Microdevices, 16 (4), 639−644. Copyright 2010 American Chemical Society.*

then flowed through the disk to isolate the wells. At that point conventional thermal cycling can be performed to achieve digital PCR and real-time detection of product based on the fluorescence in each well. The main advantage or distinction of this system is that by using a simple setup a relatively large reaction mixture is automatically distributed into nanoliter-sized volumes.

11.3 MICROFLUIDICS FOR HIGH-RESOLUTION MELTING ANALYSIS

HRMA is performed after DNA amplification is completed and is focused on the amplified region of the DNA strand (see Chapter 15). In HRMA, intercalating dyes are incorporated into the double-stranded DNA (ds-DNA). These intercalating dyes fluoresce brightly when incorporated in ds-DNA but fluoresce poorly when that ds-DNA becomes single-stranded DNA (ss-DNA), which occurs when a ds-DNA melts into ss-DNA as the temperature is increased past the melting temperature (usually from 50°C to 95°C). The melting point of ds-DNA is sensitive to DNA sequence and any mismatches, and there is a measurable drop in fluorescence as the amplicon melts. After the melting procedure, a "melt curve" is obtained by plotting fluorescence intensity as a function of temperature. The melt curve profile is unique for a particular DNA sequence (even down to a single base in the DNA sequence); hence by analyzing the melt curve profile one can identify variations in a DNA sequence.

Implementation of HRMA requires a way to change the temperature of the sample and a way to measure the fluorescence output, both of which are readily achieved on the microscale, and the thermal benefits of microfluidics apply to HRMA as well. Microfluidic systems for HRMA can be distinguished based on how the temperature gradient needed to obtain the "melt curve" is developed. There are currently two methods to develop temperature gradient in microfluidics: temporal melting and spatial melting (Crews et al., 2009).

Temporal melting is the conventional way to develop a temperature gradient in macroscale systems and has been replicated in microfluidic systems. Temporal melting is done in a fixed reservoir containing the PCR product and is basically a gradual heating process. In this case the slow heating rates are crucial for accuracy and sensitivity of the melt curve, but the slow rate of heating makes temporal melting unsuitable when fast and robust HRMA is desired. A number of microfluidic devices have been made over the past decade that use some form of temporal DNA melting for DNA identification. Most configurations involve a heating element incorporated internally into the microfluidic chip or externally outside the chip's body. The heating is done by either thermoelectricity or an external heat source such as lasers to generate a temporal thermal gradient (Lee and Fan, 2012; Athamanolap et al., 2014). Multiple images of the reservoir are taken to monitor the change in fluorescence with a change in temperature to generate a melt curve.

Spatial melting is achieved by establishing a temperature gradient across an elongated reservoir and is possible only in microfluidic systems. When the elongated reservoir is filled with a PCR product, a spatial variance of fluorescence is observed along the elongated reservoir. A single image of the reservoir is taken and the melt curve is generated from it. In spatial melting the melting can be performed on either flowing or stationary fluids, because it does not depend on time and is best suited for fast HRMA. We have reviewed significant work reported in the literature on microfluidic HRMA using spatial melting (Table 11.2).

11.3.1 DNA Methylation Detection

DNA methylation, the covalent addition of a methyl group to the cytosine base in DNA, is a central epigenetic

TABLE 11.2 Publications Reporting Significant Progress in Spatial Melting of DNA for Microfluidic-Enabled High-Resolution Melting Analysis (HRMA)

Publication Title	Highlights/Comments
Product differentiation during continuous-flow thermal gradient PCR (Crews et al., 2008)	The main focus of this publication is melting analysis. The authors show how the performance of their device carries out fast PCR/HRMA compared with commercial equipment.
Glass-composite prototyping for flow PCR with in situ DNA analysis (Pješčić_ et al., 2010)	Melting analysis is not the main focus of this publication; however the authors demonstrate how PCR and HRMA can be combined and carried out on a single microfluidic chip.
Spatial DNA melting analysis for genotyping and variant scanning (Crews et al., 2009)	**1.** Spatial microfluidic HRMA is used for single-nucleotide polymorphism scanning and genotyping. **2.** HRMA is shown in a continuous-flow regime. Up to 20 samples were processed for HRMA in serial fashion in the same device with no cleaning steps in between.
Real-time damage monitoring of irradiated DNA (Pješčić et al., 2011)	The authors demonstrate real-time measurement of DNA damage resulting from radiation exposure using a microfluidic HRMA.
Genotyping from saliva with a one-step microdevice (Pješčić and Crews, 2012)	PCR and spatial HRMA are carried out on a single disposable microfluidic chip and are shown to distinguish between human male and female saliva samples.

PCR, polymerase chain reaction.

modification and has an essential role in cellular process including genome regulation, development, and disease. Microfluidics has been shown to improve the DNA methylation detection process and improve efficiencies in time and cost by making the analysis high-throughput and sensitive with small sample volumes (Paliwal et al., 2010).

One of most well-known microfluidic DNA methylation detection methods is to use a platform with an array of microfluidic channels and an array of chambers (Weisenberger et al., 2008). The approach was tested with single methylated *PITX2* molecules. After the sample was amplified in multiple PCR reaction wells, individually amplified methylated DNA molecules were visualized via probe fluorescence signals using a high-resolution charge-coupled device (CCD) camera. This method was able to detect single-molecule DNA methylation events successfully and sensitively in a small PCR reaction volume.

Another notable approach is to use methylation-specific PCR (MSP) within an arrayed microdroplet-in-oil platform. The device has nine snowflake-like functional units arranged in a circular array (Fig. 11.3). Each function unit consists of 12 open reaction chambers, which are also arranged in a circular array and connected to a sample access port through a microfluidic network. Methylation-specific primers are predeposited into reaction chambers. The device can perform 108 MSP reactions simultaneously. Each functional unit is capable of DNA methylation analysis of multiple genes with single-sample dispensing, thereby

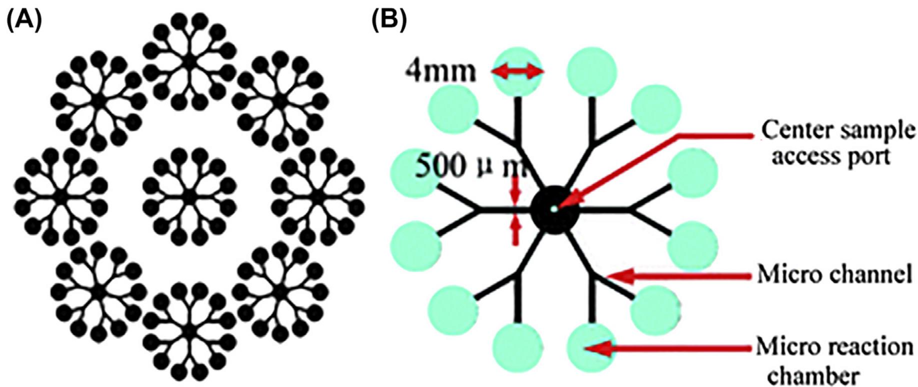


FIGURE 11.3 (A) Overall layout showing nine snowflake-like functional units arranged in a circular array. (B) Individual functional unit layout. *Reprinted with permission from Zhang, Y., et al., 2009. DNA methylation analysis on a droplet-in-oil PCR array. Lab Chip 9 (8), 1059−1064. Copyright 2009 American Chemical Society.*

significantly reducing the sample preparation time, improving throughput, and allowing for automation. This method uses mineral oil as a working fluid for actuation, which prevents contamination and evaporation of the sample. This method is exemplified by the analysis of two tumor suppressor promoters, *p15* and *TMS1* (Zhang et al., 2009).

11.3.2 Padlock and Selector Probes

Both padlock and selector probes are linear oligonucleotides with two complementary ends to the target DNA strand for DNA analysis and molecular diagnosis (Jansson, 2007) (see Chapter 18). Current molecular diagnostic approaches need manual analysis by skilled personnel, which is time-consuming and labor-intensive. The application of microfluidic technology has been increasing because of the use of a small sample volume, easy portability, and rapid detection time (Tröger et al., 2015). A few examples of microfluidic devices employing padlock and selector probe technologies are described next.

In 2005, Melin et al. developed a thermoplastic microfluidic platform for multiple purposes: sensitive detection, cell culture, and actuation. Padlock probes/rolling circle amplification (RCA) was employed in a platform for DNA detection (Melin et al., 2005). In 2006, Jarvius et al. developed an approach for quantitative single-molecular detection based on padlock probe ligation using a microfluidic system. This method was applied to sensitive detection of the bacterial pathogen *Vibrio cholerae* (Jarvius, et al., 2006).

In 2008, Mahmoudian et al. developed an integrated platform on which both circle-to-circle amplification (C2CA) and RCA were performed successfully with padlock probes. The microchip is made from poly(methyl methacrylate). A total of 25 ng bacterial genomic DNA was detected within 65 min (Mahmoudian et al., 2008). In 2010, Sato et al. created an integrated microchemical chip and combined the padlock probe and RCA on the chip. The microchip was made from glass and contains Y-shaped channels with a dam structure. A total of 88 ng *Salmonella* genomic DNA was detected using on-bead RCA on a microchip (Sato, 2010).

In 2011, Konry et al. demonstrated a droplet-based PDMS microfluidic device to detect protein markers based on padlock probe technology and RCA methods. After highly specific antigen−antibody recognition, fewer than 10 epithelial cell adhesion molecule surface tumor markers per cell were successfully detected with visual fluorescence (Konry et al., 2011). Ahlford et al. (2011) presented a microfluidic system for DNA analysis of KRAS using a highly selective padlock probes and C2CA. Tanaka et al. (2011) used a glass microchip for DNA detection based on RCA methods with padlock probes. DNA detection in small-volume samples was achieved.

In 2014, Kühnemund et al. demonstrated a digital microfluidic (DMF) chip to perform C2CA with a padlock probe. The microchip is made from glass and integrates all of the assay steps except for heating. A novel magnetic particle shuttling protocol was employed to enable high-sensitivity DNA detection (Kühnemund et al., 2014). Mezger et al. (2014) developed a rapid and sensitive microfluidic PDMS chip to detect highly variable ds-RNA viruses using padlock probes. Sato et al. developed an automated microfluidic system using RCA methods to simplify the single DNA counting process in a cell (Kuroda et al., 2014).

In some cases in which microfluidic automation is not ideal, the integration of samples can be a complicated task. Besides the potential manual steps involved, the application of padlock and selector probes technologies on a chip has largely improved DNA detection and analysis efficiency with small-volume samples. Because of the high sensitivity and reduced sample requirements, this technology can be a powerful tool in diverse diagnostic fields.

11.4 MICROFLUIDICS IN CYTOGENETICS

In the field of cytogenetics, various techniques such as fluorescence in situ hybridization (FISH) assays require expensive reagents, a large time commitment, and a need for experienced and well-trained technicians (Kwasny et al., 2012) (see Chapter 17). This makes such techniques less favorable in many cases even if they provide better results. LOC devices have been designed to improve the efficiency of many cytogenetic techniques, allowing for much faster and more reliable results (Kwasny et al., 2012).

11.4.1 Fluorescence In Situ Hybridization Analysis

Traditional FISH analysis involves a long and complex protocol that leads to the detection of genetic abnormalities. The probes used to visualize the presence of a DNA sequence have a high cost requiring more than $100 for individual tests (Kwasny et al., 2012). LOC devices have been designed that make this process much more efficient by decreasing the time commitment and the amount of probes and sample needed, and by automating the process to a greater degree, thereby achieving consistent results. Perez-Toralla et al. showed that their device was capable of decreasing the volume of sample and probes needed for FISH by a factor of 10 while cutting the time required in half. This device could be fully automated and obtain the same quality results as a traditional protocol (Perez-Toralla et al., 2015). Other devices have been able to demonstrate similar improvements, allowing up to 96 samples to be analyzed simultaneously using the same volume of probe that would usually be used for one test (Kwasny et al.,

2012). In general, LOC devices used in FISH analysis can reduce the amount of time invested and cost of reagents used, and automate processes that would otherwise require extensive training and experience. However, some of the most efficient devices are complicated and expensive to manufacture, which may limit the benefits of reducing the cost of reagents.

11.5 MICROFLUIDICS FOR PROTEIN DETECTION AND ANALYSIS

Human blood plasma contains an enormous amount of proteins, numbering around 10^{10} (Jacobs et al., 2005). Coupled with research, this demonstrates that blood plasma also plays host to critical disease biomarkers such as exosomes (Kalra et al., 2013) and micro-RNA (Mitchell et al., 2008), and paints a picture of a highly complex sample. Detection systems that can identify and analyze proteins from such mixtures effectively and rapidly can greatly enhance molecular diagnostic capabilities with downstream benefits for personalized health care (Gonzalez de Castro et al., 2013).

A variety of microfluidic systems have been developed from existing macroscale techniques for the identification and analysis of proteins. Typically, traditional protein identification approaches are derived from one of two popular methods, immunoassays or immunoblotting. Immunoassays are based on the interaction of antibodies, whether adsorbed to a surface or in free solution with specific antigens (Ng et al., 2010). Immunoblotting, on the other hand, is used first to determine the molecular mass of the protein via electrophoresis gel migration before incubation with antibodies for detection and identification (Hughes and Herr, 2012). Both methods have diffusional limitations, excessive reagent consumption, reproducibility, and throughput restrictions (Jin and Kennedy, 2015). Microfluidics can reduce diffusional distances by increasing surface area to volume ratios, reducing reagent consumption through micro- and nanofabricated channels and chambers, and automating all steps of the process (Ng et al., 2010).

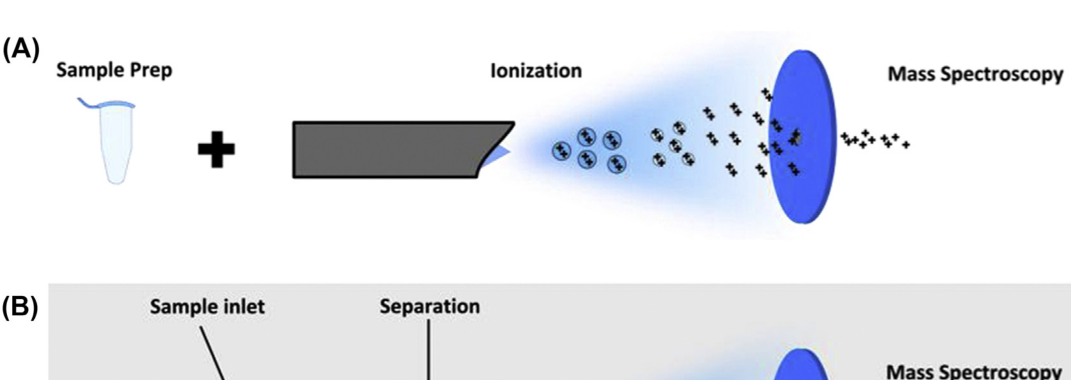
Traditional methods of protein analysis typically involve implementing two strategies for the analysis and sequencing of proteins based on mass spectrometry (MS): matrix-assisted laser deposition and ionization (MALDI) and electrospray ionization (ESI) (Domon and Aebersold, 2006). ESI uses a small nozzle or a capillary to drive reagents into the mass spectrometer for analysis (Fig. 11.4A) by reducing the charged droplets into molecular ions (Han and Gross, 2005). Alternatively, MALDI uses a dry crystalline matrix to affix a sample for laser interrogation (Fig. 11.5A). The crystalline matrix helps to desorb and ionize the sample, resulting in the sublimation of the protein species and leading to the formation of charged ions (Aebersold and Mann, 2003; Hardouin, 2007), which can then be analyzed by MS. One of the main challenges of using ESI-MS is the suppression of ions owing to high salt concentrations in the buffer (Gao et al., 2013). This makes analysis difficult or nearly impossible. Furthermore, whereas MALDI-MS is capable of analyzing high salt buffers, the structural matrix makes analysis of low mass structures difficult because of noise generation in the resulting spectra of the sample (Gao et al., 2013).

The associated challenges with using MS and the preparatory techniques can be readily addressed with microfluidics. Miniaturization leads to smaller sample volumes, reduced diffusional distances, the ability to carry out high-throughput analysis, system automation, parallelization, and process streamlining of processes (Chao and Hansmeier, 2013).

11.5.1 Overview of Electrospray Ionization–Mass Spectrometry Integrated Microfluidic Platforms

Several methods have been proposed for chip fabrication for downstream integration with MS, including traditional wet etching techniques (Lazar et al., 2006), surface micromachining (Xie et al., 2005), and rapid prototyping (Yin et al., 2005). The major features of the device are etched from a silicon substrate via photoresist deposition, followed by PDMS curing and bonding to either a silicon substrate or a glass slide. Electrical contacts can be added at any time during the fabrication process via evaporation or sputtering. Further complexity can be added to the microfluidic device by integrating valves, gates, and chambers to eliminate fitting, leaking, and blocking issues (Srbek, 2007).

Traditionally onsite filtration occurs through an area packed with microbeads. Loading the beads can be challenging ot the microscale. Vinet et al. demonstrated a robust method for fabricating 2D ordered arrays of silicon micropillars (Fig. 11.4B) using deep reactive ion etching on a silicon substrate for effective sample separation as an alternative to microbeads and precise nozzle fabrication (Mery et al., 2008). Using a tryptic peptide mixture, they were able to show effective separation and stable electrospray operation. In a separate study, Ketola et al. modified the surface of their structures with C-18 for reverse-phase separation or SiO_2 for normal separation. Furthermore, by integrating a silicon base with a glass cover, they were able to fabricate a 3D ESI tip while allowing for microfluidic chip operation with both laser-induced fluorescence and MS (Sainiemi et al., 2012). The system required only 10 nL of reagents and demonstrated fast separation and good sensitivity. One key limitation of this approach is the need



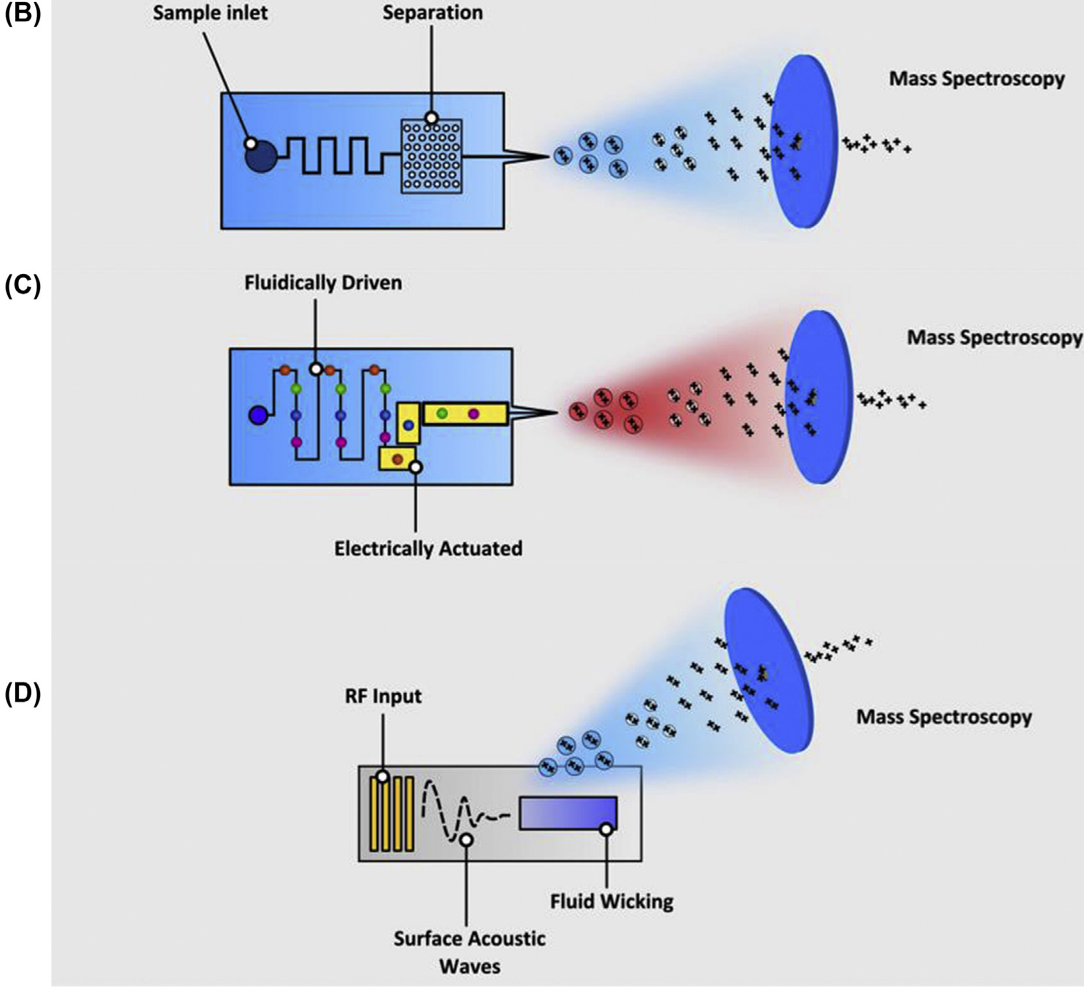


FIGURE 11.4 (A) Traditional approach to electrospray ionization (ESI)-mass spectrometry (MS). The sample is processed before ionization and MS analysis. (B) Microfluidic network in which chip separation and processing leads directly to ionization and MS analysis (Mery et al., 2008; Sainiemi et al., 2012). (C) Digital microfluidics can use either pressure-driven or electrically actuated droplet transfer to an ionization site for MS analysis (Shih et al., 2012; Baker and Roper, 2012; Ji et al., 2012). (D) Paper-based microfluidic approach coupled with surface acoustic waves for ionization and MS analysis (Ho et al., 2011). Each figure was designed to give a general idea of the methods discussed and techniques employed and is not necessarily representative of actual devices. *Prep*, preparation; *RF*, radio-frequency.

for microfabrication of the pillar arrays. Rigidity and resolution of the features are material and process dependent whereas reusability of the devices owing to clogging may also be an issue.

In an entirely different approach, Shih et al. successfully demonstrated the integration of DMF with ESI-MS analysis. An electrical potential was used to drive droplets closer to the microcapillary acting as a directly integrated ESI tip (Fig. 11.4C). Sample uptake occurred through capillary action in which an applied DC voltage was used to generate the spray interface (Shih et al., 2012). The device was used successfully to identify a specific marker in a dried blood spot sample. Similarly, Baker and Roper employed a capillary and an eductor to transfer droplets

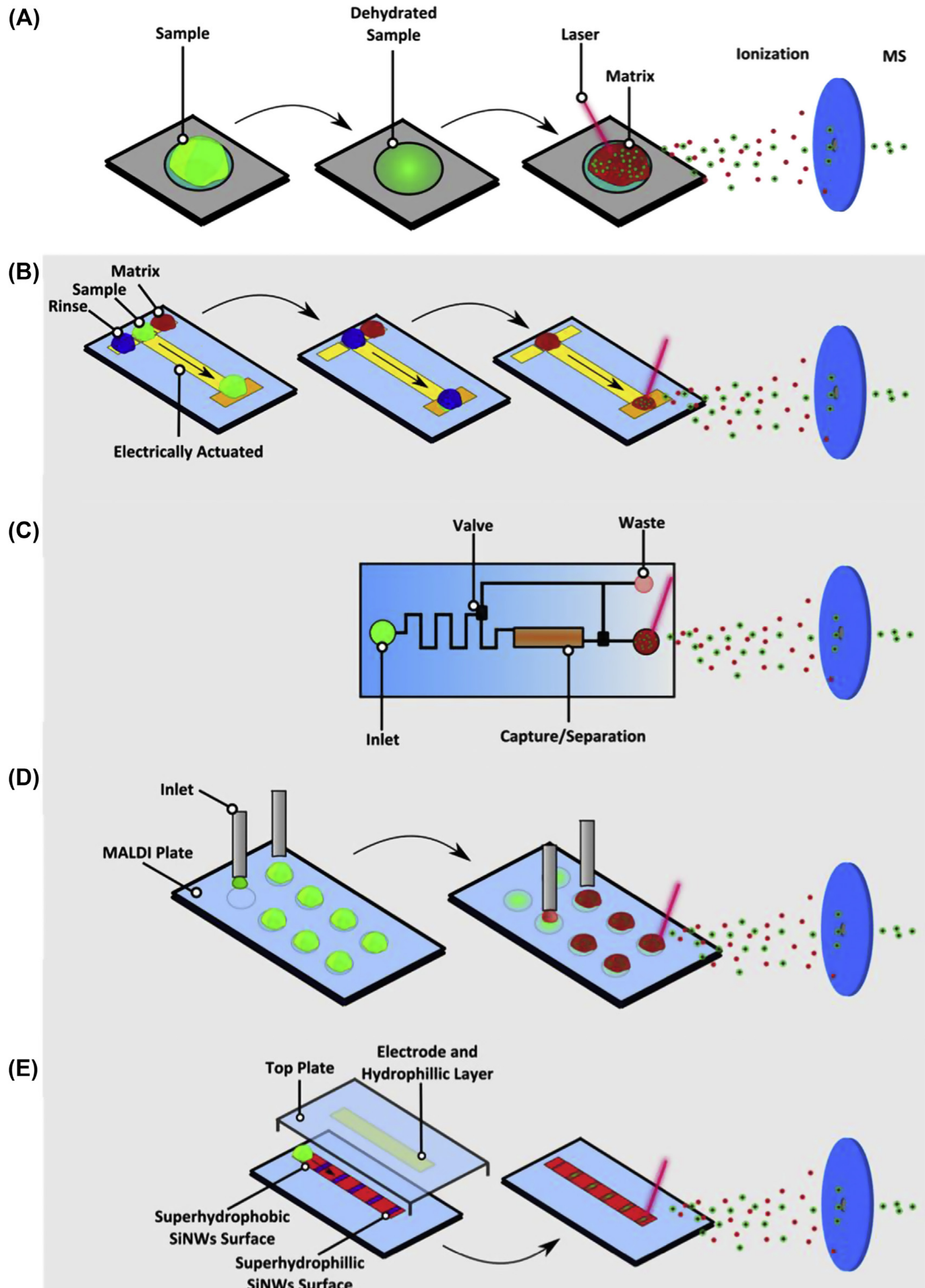


FIGURE 11.5 (A) Traditional approach to matrix-assisted laser deposition and ionization (MALDI)-mass spectrometry. A sample is added to the plate, followed by drying, matrix deposition, and laser interrogation. (B) Digital microfluidics can be used to move droplets that contain the sample, rinse, and matrix. The matrix can be deposited manually by removing the top lid (Chatterjee et al., 2010) or by moving a droplet in place (Wheeler et al., 2005). (C) Pressure (Lazar and Kabulski, 2013) or centrifugal (Thuy et al., 2010) driven flow can be used in conjunction with chromatography columns for separation, preparation, and matrix deposition. (D) Microfluidic devices can be used for contact and noncontact deposition of the target onto a MALDI plate. The matrix can be loaded before or after sample deposition (Küster et al., 2013; Ro et al., 2006). (E) Using hydrophilic capture regions, the matrix step can be removed, allowing for direct laser interrogation on the chip (Lapierre et al., 2011). Each figure was designed to give a general idea of the methods discussed and techniques employed and is not necessarily representative of actual devices. *SiNWs*, silicon nanowires.

from an open or closed setup to the ESI tip. In conjunction with nitrogen, an eductor was used to generate areas of negative pressure at the ESI tip via the Venturi effect triggering droplet movement (Baker and Roper, 2012). Continuous, high-throughput analysis of the entire droplet volume either inside the device or in ambient air is also a possibility. Furthermore, Ji et al. (2012) used DMF for rapid proteolysis (Fig. 11.4C) by encapsulating and digesting fractions inside droplets. Advantages of this approach are reduced cross-contamination, sample loss, and nonspecific absorption. One of the key limitations is in the design and fabrication of the electrical circuits to drive the process. Furthermore, as with any ESI tip, clogging can be an issue.

As mentioned earlier, a potential disadvantage of traditional approaches to microfluidic device fabrication is the need for clean room access, materials, and expertise. Paper-based microfluidic devices have a cost and are biodegradable, transportable, and effective at delivering samples to the site of interest (Mao and Huang, 2012). Ho et al. demonstrated a paper-based microfluidic device that uses wicking (Fig. 11.4D) for effective sample uptake from a reservoir for MS analysis. Surface acoustic waves were used to ionize the filtered sample at the end of the paper, effectively demonstrating the ability to process both high ionic and viscous samples, conditions that may prove difficult for traditional ESI nozzles to accommodate (Ho et al., 2011).

11.5.2 Overview of Matrix-Assisted Laser Deposition and Ionization—Mass Spectrometry—Integrated Microfluidic Platforms

Microfluidics can be integrated with the MALDI framework in several ways. The first approach manipulates a sample in a microfluidic reactor for direct deposition onto a MALDI plate. The second method uses the microfluidic device as the MALDI plate in which the sample can be prepared onsite and inserted directly into the MALDI-MS instrument for analysis (DeVoe and Lee, 2006; Lee et al., 2009).

DMF is a popular platform for integration with MALDI-MS. Chatterjee and co-workers demonstrated a microfabricated device capable of efficiently processing proteins via droplet manipulation (Fig. 11.5B) and drying, ready for MALDI interrogation (Chatterjee et al., 2010). Disulfide reduction, alkylation, and enzymatic digestion were carried out within the device, which consisted of a removable top lid and a bottom plate with integrated electrodes. Further droplet control was demonstrated by electrowetting-on-dielectrics (EWOD) with in-line sample purification for deposition onto a stainless-steel target for MALDI analysis (Wheeler et al., 2005). In a three-stage process, the sample containing peptides and impurities was moved and dried, impurities were dissolved and removed by a second droplet, and a MALDI matrix was deposited on top via a third droplet (Fig. 11.5B).

Furthermore, on-chip MALDI processing was effectively demonstrated by Lazar and Kabulski. Electro-osmotic pumps were integrated with a liquid chromatography channel for sample separation before analysis (Lazar and Kabulski, 2013). Valves were used to control the flow of sample through the slurry-loaded separation channel (Fig. 11.5C). The sample was prepared by manual addition of the MALDI matrix for analysis, successfully displaying femtomolar sensitivity for bovine cytochrome C and hemoglobin. Thuy and co-workers followed a different approach in which a CD based microfluidic device was used to prepare, digest, and analyze a sample, all in a single automated run (Thuy et al., 2010). Centrifugal force was used to drive the sample through an affinity column (Fig. 11.5C) in which the protein was captured and then tryptically digested. The digest was then eluted and co-crystallized with a MALDI matrix in 1 of the 54 designated reservoirs. Whereas droplet manipulation and in-line processing within the confines of the device offers several degrees of control over the process parameters, the main issue is still the need for a matrix that can lead to reduced signal-to-noise ratios.

Off-chip applications are also effective. Küster et al. (2013) developed a T-junction microfluidic device (Fig. 11.5D) that generated nanoliter droplets guided into a capillary for deposition onto a MALDI matrix-covered plate. The advantage of this technique is the high-throughput droplet generation potentially analyzing 26,000 droplets in a streamlined process using a detection system for automated stage movement. Furthermore, Ro and Knapp demonstrated a microfluidic device integrating an array of UV-polymerized methacrylate monolithic columns within the microfluidic channels (Fig. 11.5D) to separate tryptic-digested proteins from a peptide mixture (Ro et al., 2006). The vertically mounted device deposited droplets of the sample onto a MALDI-MS plate for analysis.

An inherent problem with using a matrix to crystallize the sample for analysis is nonhomogeneity in the formation of the crystal leading to hot spots and reduced resolution (Northen et al., 2007). Lapierre et al. (2011) demonstrated a DMF device that manipulates small amounts of volume for matrix-free laser interrogation. A small droplet containing the sample was actuated along a channel patterned with superhydrophobic and superhydrophyllic areas (Fig. 11.5E) on top of a silicon nanowire interface that captured some of the liquid. Upon drying and laser interrogation, the silicon nanowire interface acts as an inorganic target allowing MS analysis. Although it is highly sensitive and matrix free, the

key limitation of this approach is the complexity associated with the fabrication of the device.

11.5.3 Overview of Popular Microfluidic Detection Platforms

One strength of microfluidics is direct customization of most traditional immunoassay protocols. Several strategies have been proposed for enhancing limits of detection, including gold nanoparticles, which act as nanoelectrodes with high electrical conductivity and surface area for antibody attachment and detection (Mani et al., 2009). Because of their tunability, brightness, high absorption coefficients, and photostability, quantum dots have also yielded highly sensitive results (Li et al., 2010). In a different approach, Karns and Herr used electrophoretic immunoassay separation of endogenous tear protein biomarkers to obtain mobility and immunoaffinity information from 1-μL samples (Karns and Herr, 2011).

Fast, efficient, high-throughout platforms may allow for enhanced sample quantification and, as a result, better treatment strategies. Protein microarrays have contributed a great deal toward realizing this goal by using pin printing, microstamps, or microflow printing assays (Romanov, 2014). Simple, cost-effective, high-throughput microfluidic devices with high sensitivity have also been described for rapid diagnoses of human immunodeficiency virus and syphilis (Chin et al., 2011). Selecting the correct microarray platform is crucial because platforms may vary in the types of molecules that they can print, the quality of the spots, the throughput, and operational requirements.

In addition, DMF has also been employed as miniaturized immunoassay reactors. In these systems, protein detection typically relies on optical methods; however, electrochemical detection has also been demonstrated (Shamsi et al., 2014). One advantage of DMF is the generation of highly tunable droplets. Vergauwe et al. (2011) demonstrated a highly sensitive EWOD system capable of both heterogeneous and homogeneous immunoassays by droplet manipulation. An issue with using DMF for immunoassays is sample recovery. Whereas detection has been thoroughly demonstrated, recovering protein for subsequent study is still challenging.

Microfluidics has also been effectively used for western blotting. Herr's group dedicated a significant amount of effort in improving most facets of traditional western blots including completely automated western blots with reusable chips (He and Herr, 2010) and fully integrated, rapid lectin blotting through the removal of sodium dodecyl sulfate from resolved protein peaks via photopatterned microfilters within the microfluidic device (He et al., 2011). A similar approach was used for analysis of human sera and cell lysate employing a glass microfluidic chip that resulted in rapid operation on the order of 10–60 min with picomolar detection limits (Hughes and Herr, 2012). Cleaning and reusability of such devices were demonstrated but it is not clear how many times these devices may be regenerated before contamination or material degradation becomes an issue.

An electrostatic immobilization gel was developed as an alternative to the sandwich format typically used within microfluidic western blots. The result was a reduction of reagents consumption on the order of 200 times and a reduction in assay duration by 12 times achieved through charge interactions (Kim et al., 2012). As an alternative to introducing samples into a channel for separation and immobilization, Jin (2013) fabricated a microfluidic chip for direct deposition of sieve-separated protein bands on a perpendicularly mounted polyvinylidene fluoride membrane. Using this approach they demonstrated reliable, reusable, and reproducible separation and multiple injections using the same channel and capture membrane.

11.6 MICROFLUIDIC SAMPLE PREPARATION

There have been significant advances in the application of microfluidics for molecular diagnostics applications. Extensive research has gone into integrating molecular analysis systems (nucleic acids, proteins, pathogens, and small molecules) on-chip. These have led to a reduction in the costs of reagents and user interaction with the instruments. These platforms have leveraged advances in technologies such as PCR, CE, FISH, surface plasmon resonance surface enhanced Raman scattering, giant magnetoresistance, and electrical/electrochemical/mechanical detection among others (Kim et al., 2009). All of these methods typically require some form of off-chip sample preparation (SPrep). Unfortunately, advances in on-chip SPrep have been overlooked compared with downstream processes such as analysis and sensing (Brehm-Stecher et al., 2009; Mariella, 2008; Kim et al., 2009). The dependence on traditional off-chip sample pretreatment involving expensive equipment and trained personnel has prevented the translation of these advances to POC (Byrnes et al., 2015).

Sample preparation steps include cell lysis, washing, centrifugation, separation, filtration, and elution. These techniques performed using the conventional route are highly labor-intensive and time-consuming, involve multiple steps, and require expensive laboratory equipment (Byrnes et al., 2015). For instance, nucleic acid (NA) extraction involves multiple steps to collect DNA or RNA from raw samples such as whole blood, urine, and saliva. On-chip integration of these steps can reduce the total analysis time. The lower sample and regent consumption in microfluidic systems enables a lower cost of analysis. Also, an enclosed sample-in, answer-out system reduces the

chance for cross-contamination. On-chip sample integration typically involves adaptation and modification of conventional macroscale laboratory methods to fit microfluidic formats (Byrnes et al., 2015; Kim et al., 2009). This section describes some advances in microfluidic SPrep for specific molecular diagnostic techniques such as PCR and DNA sequencing. Table 11.3 summarizes some of these microfluidic platforms.

11.6.1 Microfluidic Sample Preparation for Polymerase Chain Reaction

PCR-based methods have opened up a myriad of possibilities in diagnostics for pathogens and infectious diseases in both clinical and environmental settings. Several commercial Food and Drug Administration—approved PCR platforms exist (Priye and Ugaz, 2016), but most require either manual off-chip SPrep or separate automated SPrep systems involving bench-top equipment such as a centrifuge. Hence these tests are not Clinical Laboratory Improvement Amendments waived or POC (Mitchell et al., 2012). Therefore on-chip integration of SPrep will enable true sample-in, answer-out PCR platforms for POC use. They also provide advantages such as lower reagent consumption, faster cycling times, lower cost per test, and automated processing requiring minimally trained personnel (Oblath et al., 2013). Consistency of SPrep is known to affect results of digital RT-PCR (Thompson

TABLE 11.3 Microfluidic Sample Preparation (SPrep) Platforms Used in Nucleic Acid (NA) Amplification and Next-Generation DNA Sequencing (NGS), With On-Chip SPrep Steps Listed

Methods	Steps Incorporated On-Chip	Notes
PCR, RT-PCR (Kim et al., 2010)	NA extraction (SPE) using nanoporous AOM and amplification	Inhibitory role of AOM during PCR, lower retention of RNA in AOM
RT-PCR (Oblath et al., 2013)	NA extraction using nanoporous AOM and amplification	Off-chip thermal lysis, inhibitory effects of AOM minimized by adding bovine serum albumin, and additional *Taq* polymerase to master mix
RT-PCR (Czilwik et al., 2015)	Chemical lysis, NA extraction (silica-coated magnetic beads), consensus multiplex PCR preamplification, and geometrically multiplexed RT-PCR	Demonstrated specific detection of four model pathogens down to 3 colony-forming units in serum. System incorporated prestored reagents but requires prior serum separation from whole blood, which might impede POC use
Multiplex array PCR, DEP (Cai et al., 2014)	Pathogen capture (DEP), thermal lysis, and multiplex array PCR	Simultaneous detection of three pathogens in 3 h. Preloaded PCR reagents and sample-in, answer-out capability for potential POC use, but system relied on conventional thermal cycler, power supply (for DEP), and fluorescence microscope, potentially impacting its portability for POC use
HDA (Huang et al., 2013)	Chemical lysis, NA extraction (SPE), and HDA based amplification	Lysis and NA extraction were performed on system (SNAP) distinct from NA amplification. Electricity-free NA amplification system consisted of a cyclic olefin polymer µChip placed inside a Styrofoam cup with commercially available toe warmers acting as heaters. Extracted NA was manually transferred over to µChip for amplification
RPA (Kim et al., 2014)	Thermal lysis, RPA-based amplification, and visual detection using lateral flow strips	Centrifugal platform for food-borne pathogen detection. Laser diode was used for actuation of Ferro-Wax valves, thermal lysis, and amplification
Real-time LAMP (Sun et al., 2015)	Chemical lysis, NA extraction (magnetic beads), and isothermal amplification	Integration of SPrep to LAMP-based detection on same eight-chamber thermoplastic chip
Convective PCR (Priye et al., 2016)	Chemical lysis, NA extraction (SPE), PCR amplification, and fluorescence detection	Portable convective thermocycler loaded on drone with SPrep leveraging drone's motors Time-resolved fluorescence detection and quantification using integrated smartphone camera. Sample lysis and loading are performed manually using the platform
NGS Library prep (Kim et al., 2013)	NGS library prep	DMF platform to prepare NGS libraries from a few nanograms of genomic DNA. DNA extraction and purification performed off-chip

AOM, aluminum oxide membrane; *DEP*, dielectrophoresis; *DMF*, digital microfluidics; *HDA*, helicase-dependent amplification; *LAMP*, loop-mediated isothermal amplification; *PCR*, polymerase chain reaction; *POC*, point-of-care; *RPA*, recombinase polymerase amplification; *RT*, reverse-transcribed; *SNAP*, portable system for nucleic acid preparation; *SPE*, solid-phase extraction; *µChip*, microfluidic chip.

et al., 2014), so these advances are critical. Also, microfluidic chips (μChips) can be disposable, eliminating contamination between samples, which is important because the high sensitivity provided by PCR poses issues owing to NA contamination. This section focuses on advances in microfluidic devices with integrated SPrep-PCR and integrated SPrep−PCR detection capability. For a detailed review of microfluidic PCR and similar integrated systems before 2013, the reader can refer to Park et al. (2014).

Kim et al. (2010) developed a μChip that integrates solid-phase extraction and amplification of NAs into a single reaction chamber. A nanoporous aluminum oxide membrane (AOM) was employed for the solid-phase extraction (SPE) of NAs. A μChip integrated DNA extraction using monolithic AOM and seven parallel reaction wells for real-time amplification of extracted DNAs. The system demonstrated the detection of bacterial pathogens in a whole saliva sample (Oblath et al., 2013). A disposable microfluidic chip with integrated SPE for NA extraction and RT-PCR was used to amplify influenza An RNA in human nasopharyngeal aspirate and nasopharyngeal swab specimens (Mitchell et al., 2012). However, the PCR products were detected off-chip by CE. In all of these devices, sample lysis was performed off-chip.

A μChip integrating electrochemical cell lysis, PCR, CE-based separation, and amperometric detection was reported for detection of pathogens (Jha et al., 2012). However, the system displayed shortcomings associated with temperature control in PCR reactions (Adley, 2014). Czilwik et al. reported a centrifugal microfluidic-based platform (LabDisk) using prestored reagents with integrated DNA extraction, consensus multiplex PCR preamplification, and geometrically multiplexed, species-specific RT-PCR. The system was able to detect low concentrations of pathogens [2 colony-forming units (CFU)/200 μL] from serum samples (Czilwik et al., 2015). The system requires serum separation from whole blood off-chip. However, serum or plasma separation has been demonstrated in microfluidic formats and could be integrated with the LabDisk system. Cai et al. (2014) demonstrated a completely integrated microfluidic device fabricated using "SlipChip" technologies for the detection of pathogens in biological samples (blood). The platform employs dielectrophoresis for extraction, multiplex array PCR for amplification, and end-point fluorescence for the simultaneous detection of three different pathogens. However, the limit of detection of 10^3 CFU/mL reported using the platform is low and impedes its potential use in practical applications.

11.6.2 Microfluidic Sample Preparation for Isothermal Amplification

Isothermal amplification uses a single temperature as opposed to cycling between multiple temperatures as in the case of PCR. Because no thermal cycling is involved, there is generally a reduced need for power, especially over long-term use, which makes it suitable for POC (Almassian and Nelson, 2013).

Huang et al. applied helicase-dependent isothermal amplification for detection of Clostridium *difficile* in stool samples. The electricity-free system consists of a μChip in a Styrofoam cup (the insulator) able to maintain its temperature at $65 \pm 2°C$. SPrep employed a stand-alone, pressure-driven "portable system for nucleic acid preparation" (SNAP), powered by a bicycle pump. It consisted of four subsystems: a sample input and mixer, a fluid-buffering coil, an air pressure accumulator, and a sample extraction cartridge. The sample lysis and NA extraction were performed using the SNAP, which was distinct from the amplification system, requiring manual transfer of extracted NA. The downstream detection of amplicons was also performed off-chip (Huang et al., 2013). Hence, integration of sample lysis and NA extraction along with readout for amplicon detection would be necessary to achieve sample-in, answer-out capability. A device consisting of flexible plastic substrate containing chambers in a reel-to-reel cassette format was used for loop-mediated isothermal amplification (LAMP) and colorimetric detection. The system performs thermal shock lysis of hard-to-lyse, gram-positive bacteria on-chip, but fluid/reagent metering and mixing is done manually (pipetting), which potentially limits POC use (Safavieh et al., 2014). Kim et al. developed a centrifugal microfluidic device integrating DNA extraction, isothermal recombinase polymerase amplification (RPA), and detection onto a single disc (Fig. 11.6). A laser diode was used for wireless control of valve actuation, cell lysis, and noncontact heating during the RPA step. Immunomagnetic separation has proven to be an effective tool for preconcentration of pathogens from large-volume samples containing potential interferents (Jayamohan et al., 2015). However, for this device the immunomagnetic extraction of pathogens was performed off-chip (Kim et al., 2014). Similar centrifugal platforms have been applied to SPrep for RT-PCR (viral detection) (Stumpf et al., 2016), LAMP (Sayad et al., 2016), and digital PCR (Burger et al., 2016; Schuler et al., 2016). An eight-chamber LOC device integrating cell lysis, immunomagnetic bead-based DNA extraction, LAMP, and fluorescence detection was reported (Sun et al., 2015). The system was reported to have true sample-in, answer-out capability for the detection of *Salmonella*.

11.6.3 Microfluidic Sample Preparation for Sequencing

DNA sequencing refers to the process of determining the precise order of nucleotides within a DNA strand. Next-generation DNA sequencing (NGS) broadly refers to advances in sequencing that have enabled high-throughput, inexpensive, rapid, whole-genome sequencing (Metzker, 2010) (see Chapters 22 and 23). NGS has been widely

FIGURE 11.6 Lab-on-a-disc system for the detection of food-borne pathogens. (A) The system consisted of two polycarbonate layers with integrated strip sensors and metal heaters. (B) The disc consists of chambers for cell lysis, isothermal amplification, metering, dilution, and detection. (C) Schematic of the setup showing computer-controlled spinning motor, laser for the thermal actuation of Ferro-Wax valves and isothermal DNA amplification, and a charge-coupled device (CCD) camera and strobe light to visualize the rotating disc in real time. *Reprinted with permission from Kim, T.-H., Park, J., Kim, C.-J., Cho, Y.-K., 2014. Fully integrated lab-on-a-disc for nucleic acid analysis of food-borne pathogens. Anal. Chem., 86 (8), 3841–3848. Copyright 2014 American Chemical Society.*

applied to elucidate genetic information for applications such as pathogen discovery and the identification of genetic abnormalities associated with human disease (Kim et al., 2013). Genome sequencing has come a long way since the conclusion of the Human Genome Project. Sequencing platforms have evolved from bulky systems (860 kg PacBios RSII) to relatively inexpensive, pocket-sized versions (Oxford MinION and SmidgION) (Erlich, 2015; Pennisi, 2016). Advances leveraging nanopores have the potential to democratize sequencing (Quick et al., 2016). The cost of sequencing an individual full genome has plummeted from USD $2.7 billion (Human Genome Project) to the current $1000, outpacing even Moore's law (Hayden, 2014). As in other areas of molecular diagnostics, automated DNA SPrep is a key challenge in achieving a small-footprint, sample-in, data-out sequencing platform (Hayden, 2014; Coupland, 2010). For instance, advances in the area of automation of preparation methods for formatting sample DNA into sequencing-ready libraries have lagged behind significant advances in NGS (Kim et al., 2013). However, advances have the potential to narrow the gap. This section will focus on published work on downstream microfluidic SPrep for genome sequencing (library preparation). We have published a detailed review of upstream microfluidic DNA SPrep techniques (cell lysis and DNA extraction), to which the reader can refer (Kim et al., 2009).

Patel and co-workers developed a DMF platform as a fluid distribution hub (Fig. 11.7). The platform enables the integration of multiple subsystem modules into an automated NGS library SPrep system. The central DMF hub is

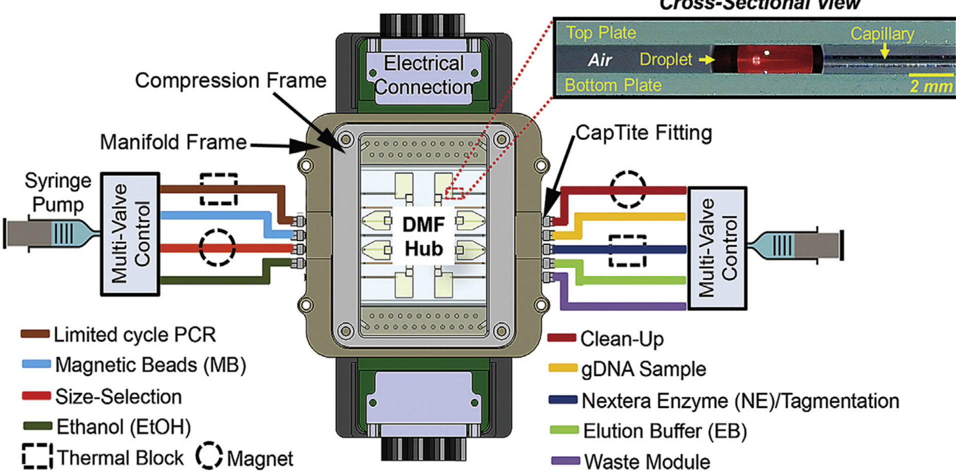


FIGURE 11.7 Digital microfluidics system for preparing DNA libraries for sequencing. The system integrated multiple reagent and sample preparation modules (depicted in different colors), magnets (for magnetic bead–based separation/cleanup), and thermal blocks (for thermal cycling) coupled to module tubing (for sample preparation) and multivalve syringe pumps (for liquid handling). *gDNA*, genomic DNA; *PCR*, polymerase chain reaction.

interfaced through novel capillary interconnects to external fluidic modules for highly repeatable transfer of liquid (Hanyoup et al., 2011). The authors used a similar DMF platform to prepare sequencer-ready DNA libraries for analysis by the Illumina MiSeq sequencer (Kim et al., 2013). Cell lysis and DNA extraction steps were performed off-chip using conventional laboratory methods. Tan and co-workers reported a novel microfluidic device capable of performing an arbitrary number of serial reaction-purification steps on 16 independent samples. They applied the device to implement protocols to generate NGS libraries from bacterial and human genomic DNA samples. Similar DMF-based platforms (VolTRAX) are in the process of being commercialized for point-of-use automated sample preparation (Dodsworth, 2015; Oxford Nanopore, 2016).

A single cell sequencing method (Drop-Seq) employing droplet microfluidics was proposed by Macosko et al. (2015). The system encapsulates one cell per emulsion droplet, lyses them, and uniquely barcodes the RNA of each cell using DNA-barcoded microgel beads. Hence, a number of conventional processing steps are compressed into a single step, creating a scalable method for in situ library preparation (Erlich, 2015). Klein et al. (2015) reported a similar technique (inDrop) for barcoding the RNA from thousands of individual cells. The technique was used to probe transcriptional variability in mouse embryonic stem cells. These methods had limitations seen in droplet microfluidics, such as variability in the number of cells per droplet (dictated by Poisson statistics). GnuBIO (a Bio-Rad company) is commercializing a microfluidic-based system for genomic library preparation, integrated within their bench-top sequencer. The system promises to be a true “sample in, answer out” DNA sequencing solution (Erlich, 2015; Business-Wire, 2014). Genetic analysis of minute amounts of DNA and RNA at the level of a single cell using NGS methods is increasingly being relied upon to understand biological complexity previously concealed when employing conventional techniques (Thompson et al., 2014). These involve whole-genome amplification (WGA) or reverse transcription and WGA before NGS. SPrep improvements have helped improve the accuracy of the reverse transcription and preamplification steps. Still, current RNA-sequencing methods cannot be considered as absolute counting technologies (Thompson et al., 2014). Wu et al. compared the sensitivity and reproducibility of single-cell whole transcriptome preparations. They reported less gene dropout and improved reproducibility and accuracy by performing reverse transcription and preamplification steps in microfluidic volumes of the C1 device (Fluidigm) rather than tube-based preparations (Thompson et al., 2014).

Despite significant advances in applying microfluidics in SPrep for sequencing, integrated sample-in, sequence-out platforms are lacking. This is especially critical because real-time, portable sequencers are being developed (Quick et al., 2016) and manual SPrep remains a critical bottleneck preventing their widespread use.

11.6.4 Microfluidic Sample Preparation in Cytogenetics

Often in cytogenetic studies, the cell samples obtained are small and difficult to work with. In these cases, LOC devices have been designed that allow for culturing of many different kinds of cells. These devices often allow for greater efficiency when working with small volumes, as well as greater ease of operation and automation, which decreases the risk of human error (Tehranirokh et al., 2013). SPrep is not limited to culturing; it includes other processes that are necessary for FISH assays, karyotyping, or other cytogenetic testing. Shah et al. built a device that integrates multiple stages of the process, allowing for initial culturing, arrest, and fixation of metaphase cells. It also has the ability to prepare metaphase chromosome spreads on glass slides for metaphase FISH analysis (Kwasny et al., 2012). Creating chromosome spreads has been described as more of an art than a science, but many devices (including the device built by Shah et al.) are being built that make the creation of these spreads more reliable and repeatable (Kwasny et al., 2012, 2014).

11.7 MICROFLUIDICS IN CELL SORTING

Cell sorting can be performed using both electrokinetically and hydrodynamically driven mechanisms in microfluidic devices (Paegel et al., 2003). Fu et al.(2002) demonstrated single-cell handling feasibility. Grover et al. (2003) presented a hybrid glass-PDMS microfluidic device with elastomer valves and pumps, which provided reliable fluid control of chips . In 2010, Gagnon et al. developed a closed-loop microfluidic device for yeast cell separation using AC electrokinetic components. In 2011, a microfluidic device used for high-efficiency circulating tumor cell selection was presented (Dharmasiri et al., 2011). Karimi et al. (2013) reported cell focusing and sorting using hydrodynamic mechanisms. More recently, Song et al. (2015) developed an electrokinetic microfluidic device for cell sorting.

11.8 FUTURE OF MICROFLUIDICS FOR MEDICAL DIAGNOSTICS

The medical diagnostics field is rapidly being transformed by the introduction and optimization of microfluidic devices. Because of the obvious size match between microfluidics and most biological processes at the cellular and subcellular level, the use of microfluidics will naturally continue to be applied to medical diagnostics. Complexity is much more readily introduced at the microscale than in traditional

formats, so it is anticipated that increasingly complex microfluidic devices able to perform multiple diagnostic processes at the same time will be developed. Already systems can detect multiple diseases or pathogens simultaneously. These devices currently rely on the same methods on multiple targets. As sample preparation techniques improve, though, it is anticipated that multiple processes will be possible on one chip, allowing the detection of DNA, proteins, chemicals, and other biomolecules on one device. As multiplexed protein detection joins multiplexed DNA detection and analysis, small-scale, portable, microfluidic-based instruments will be able to perform massively parallel analysis simultaneously at increasingly lower costs, making the need for individual tests less relevant, especially when physicians are attempting to make a diagnosis with symptoms suggestive of any of several possible conditions. Thus, the dream of a single instrument to perform nearly any molecular diagnostic procedure can be seen now and will likely occur in the next 10–20 years.

Even further along may be the opportunity for combinations of microfluidics and nucleic acid sequencing technologies. Sequencing is becoming ubiquitous, with prices falling rapidly. As sequencing becomes a commodity and can be completed more rapidly and instruments are becoming further miniaturized, sample preparation using microfluidic instruments will become the limiting factor in developing "near universal" molecular diagnostic tools, at least for anything that can be diagnosed using a nucleic acid sequence. Of course, biological understanding and computer software will need to keep up, but it is clear that both will move quickly if the opportunity is available. Thus, universal sample preparation instruments for high-throughput sequencing may be one of the major challenges for microfluidics in the next decade and for the foreseeable future of molecular diagnostics. Rapid development in this area will only be possible as more robust, generic, and complex microfluidics allow, which makes this a potent area for high-impact research activities.

REFERENCES

Abate, A.R., et al., 2013. DNA sequence analysis with droplet-based microfluidics. Lab Chip 13 (24), 4864–4869.

Adley, C.C., 2014. Past, present and future of sensors in food production. Foods 3 (3), 491–510.

Aebersold, R., Mann, M., 2003. Mass spectrometry-based proteomics. Nature 422 (6928), 198–207.

Ahlford, A., et al., 2011. A Microfluidic Platform for Personalized Cancer Diagnostics by Padlock Probe Ligation and Circle-to-Circle Amplification. s.n, Seattle, pp. 61–63.

Almassian, D.R., Cockrell, L.M., Nelson, W.M., 2013. Portable nucleic acid thermocyclers. Chem. Soc. Rev. 42 (22), 8769–8798.

Almoguera, C., et al., 1988. Most human carcinomas of the exocrine pancreas contain mutant c-K-*ras* genes. Cell 53 (4), 549–554.

Asiello, P.J., Baeumner, A.J., 2011. Miniaturized isothermal nucleic acid amplification, a review. Lab Chip 11 (8), 1420–1430.

Athamanolap, P., Shin, D.J., Wang, T.-H., 2014. Droplet array platform for high-resolution melt analysis of DNA methylation density. J. Lab. Autom. 19 (3), 304–312.

Baker, C.A., Roper, M.G., 2012. Online coupling of digital microfluidic devices with mass spectrometry detection using an eductor with electrospray ionization. Anal. Chem. 84 (6), 2955–2960.

Becker, H., 2010. Mind the gap! Lab Chip 271–273.

Blazej, R.G., Kumaresan, P., Mathies, R.A., 2006. Microfabricated bioprocessor for integrated nanoliter-scale Sanger DNA sequencing. Proc. Natl. Acad. Sci. U. S. A 103 (19), 7240–7245.

Brehm-Stecher, B., Young, C., Jaykus, L.-A., Tortorello, M.L., 2009. Sample preparation: the forgotten beginning. J. Food Prot. 72 (8), 1774–1789.

Burger, R., Amato, L., Boisen, A., 2016. Detection methods for centrifugal microfluidic platforms. Biosens. Bioelectron. 76, 54–67.

Business-Wire, 2014. Bio-rad Acquires GnuBIO, Developer of Droplet-based DNA Sequencing Technology (Online) Available at: http://gnubio.com/news/bio-rad-acquires-gnubio-developer-of-droplet-based-dna-sequencing-technology-2/.

Byrnes, S.A., et al., 2015. One-step purification and concentration of DNA in porous membranes for point-of-care applications. Lab Chip.

Cai, D., et al., 2014. An integrated microfluidic device utilizing dielectrophoresis and multiplex array PCR for point-of-care detection of pathogens. Lab Chip 14 (20), 3917–3924.

Cao, W., et al., 2015. Automated microfluidic platform for serial polymerase chain reaction and high-resolution melting analysis. J. Lab. Autom. http://dx.doi.org/10.1177/2211068215579015.

Chang, C.-M., et al., 2013. Nucleic acid amplification using microfluidic systems. Lab Chip 13 (7), 1225–1242.

Chao, T., Hansmeier, N., 2013. Microfluidic devices for high-throughput proteome analyses. Proteomics 13 (3–4), 467–479.

Chatterjee, D., et al., 2010. Integration of protein processing steps on a droplet microfluidics platform for MALDI-MS analysis. Anal. Chem. 82 (5), 2095–2101.

Chin, C.D., et al., 2011. Microfluidics-based diagnostics of infectious diseases in the developing world. Nat. Med. 17 (8), 1015–1019.

Chin, C.D., Linder, V., Sia, S.K., 2012. Commercialization of microfluidic point-of-care diagnostic devices. Lab Chip 12 (12), 2118–2134.

Coupland, P., 2010. Microfluidics for the upstream pipeline of DNA sequencing – a worthy application? Lab Chip 10 (5), 544–547.

Craw, P., Balachandran, W., 2012. Isothermal nucleic acid amplification technologies for point-of-care diagnostics: a critical review. Lab Chip 12 (14), 2469–2486.

Crews, N., Wittwer, C., Palais, R., Gale, B., 2008. Product differentiation during continuous-flow thermal gradient PCR. Lab Chip 8 (6), 919–924.

Crews, N., et al., 2009. Spatial DNA melting analysis for genotyping and variant scanning. Anal. Chem. 81 (6), 2053–2058.

Czilwik, G., et al., 2015. Rapid and fully automated bacterial pathogen detection on a centrifugal-microfluidic LabDisk using highly sensitive nested PCR with integrated sample preparation. Lab Chip 15 (18), 3749–3759.

Derveaux, S., et al., 2008. Synergism between particle-based multiplexing and microfluidics technologies may bring diagnostics closer to the patient. Anal. Bioanal. Chem. 391 (7), 2453–2467.

DeVoe, D.L., Lee, C.S., 2006. Microfluidic technologies for MALDI-MS in proteomics. Electrophoresis 27 (18), 3559–3568.

Dharmasiri, U., et al., 2011. High-throughput selection, enumeration, electrokinetic manipulation, and molecular profiling of low-abundance circulating tumor cells using a microfluidic system. Anal. Chem. 83 (6), 2301−2309.

Dodsworth, S., 2015. Genome skimming for next-generation biodiversity analysis. Trends Plant Sci. 20 (9), 525−527.

Domon, B., Aebersold, R., 2006. Mass spectrometry and protein analysis. Science 312 (5771), 212−217.

Duffy, D.C.J., McDonald, C., Schueller, O.J., Whitesides, G.M., 1998. Rapid prototyping of microfluidic systems in poly(dimethylsiloxane). Anal. Chem. 4974−4984.

Erlich, Y., 2015. A vision for ubiquitous sequencing. bioRxiv, 019018.

Farrar, J.S., Wittwer, C.T., 2015. Extreme PCR: efficient and specific DNA amplification in 15-60 seconds. Clin. Chem. 61, 145−153.

Fredrickson, C.K., Fan, Z.H., 2004. Macro-to-micro interfaces for microfluidic devices. Lab Chip 526−533.

Fu, A.Y., et al., 2002. An integrated microfabricated cell sorter. Anal. Chem. 74 (11), 2451−2457.

Gagnon, Z., Mazur, J., Chang, H.-C., 2010. Integrated AC electrokinetic cell separation in a closed-loop device. Lab Chip 10 (6), 718−726.

Gao, D., Liu, H., Jiang, Y., Lin, J.-M., 2013. Recent advances in microfluidics combined with mass spectrometry: technologies and applications. Lab Chip 13 (17), 3309−3322.

Gonzalez de Castro, D., Clarke, P.A., Al-Lazikani, B., Workman, P., 2013. Personalized cancer medicine: molecular diagnostics, predictive biomarkers, and drug resistance. Clin. Pharmacol. Ther. 93 (3), 252−259.

Grover, W.H., et al., 2003. Monolithic membrane valves and diaphragm pumps for practical large-scale integration into glass microfluidic devices. Sens. Actuators B Chem. 89 (3), 315−323.

Hadd, A.G., Goard, M.P., Rank, D.R., Jovanovich, S.B., 2000. Submicroliter DNA sequencing for capillary array electrophoresis. J. Chromatogr. A 894 (1), 191−201.

Han, X., Gross, R.W., 2005. Shotgun lipidomics: electrospray ionization mass spectrometric analysis and quantitation of cellular lipidomes directly from crude extracts of biological samples. Mass Spectrom. Rev. 24 (3), 367−412.

Hanyoup, K., et al., 2011. Automated digital microfluidic sample preparation for next-generation DNA sequencing. J. Lab. Autom. 16 (6), 405−414.

Hardouin, J., 2007. Protein sequence information by matrix-assisted laser desorption/ionization in-source decay mass spectrometry. Mass Spectrom. Rev. 26 (5), 672−682.

Hauck, T.S., Giri, S., Gao, Y., Chan, W.C., 2010. Nanotechnology diagnostics for infectious diseases prevalent in developing countries. Adv. Drug Deliv. Rev. 62 (4), 438−448.

Hayden, E.C., 2014. Technology: the $1,000 genome. Nat. News 507 (7492), 294−295.

He, M., Herr, A.E., 2010. Automated microfluidic protein immunoblotting. Nat. Protoc. 5 (11), 1844−1856.

He, M., Novak, J., Julian, B.A., Herr, A.E., 2011. Membrane-assisted online renaturation for automated microfluidic lectin blotting. J. Am. Chem. Soc. 133 (49), 19610−19613.

He, Y., Pang, H.-M., Yeung, E.S., 2000. Integrated electroosmotically-driven on-line sample purification system for nanoliter DNA sequencing by capillary electrophoresis. J. Chromatogr. A 894 (1), 179−190.

Hindson, B.J., Ness, K.D., et al., 2011. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. Anal. Chem. 83 (22), 8604−8610.

Ho, J., et al., 2011. Paper-based microfluidic surface acoustic wave sample delivery and ionization source for rapid and sensitive ambient mass spectrometry. Anal. Chem. 83 (9), 3260−3266.

Hsieh, A.T.-H., Pan, P.J.-H., Lee, A.P., 2009. Rapid label-free DNA analysis in picoliter microfluidic droplets using FRET probes. Microfluid. Nanofluidics 6 (3), 391−401.

Huang, S., et al., 2013. Low cost extraction and isothermal amplification of DNA for infectious diarrhea diagnosis. PLoS One 8 (3).

Hughes, A.J., Herr, A.E., 2012. Microfluidic western blotting. Proc. Natl. Acad. Sci. U. S. A 109 (52), 21450−21455.

Jacobs, J.M., et al., 2005. Utilizing human blood plasma for proteomic biomarker discovery. J. Proteome Res. 4 (4), 1073−1085.

Jansson, R., 2007. Development of a Solid-phase Padlock Probe Technology Using Microfluidics. Uppsala University School of Engineering, Uppsala.

Jarvius, J., et al., 2006. Digital quantification using amplified single-molecule detection. Nat. Methods 3 (9), 725−727.

Jayamohan, H., et al., 2015. Highly sensitive bacteria quantification using immunomagnetic separation and electrochemical detection of guanine-labeled secondary beads. Sensors 15 (5), 12034−12052.

Jayamohan, H., Sant, H.J., Gale, B.K., 2013. Applications of microfluidics for molecular diagnostics. In: Microfluidic Diagnostics. Springer, s.l., pp. 305−334

Jenkins, G., Mansfield, C.D., 2013. Microfluidic Diagnostics: Methods and Protocols. Humana Press, s.l.

Jha, S.K., et al., 2012. An integrated PCR microfluidic chip incorporating aseptic electrochemical cell lysis and capillary electrophoresis amperometric DNA detection for rapid and quantitative genetic analysis. Lab Chip 4455−4464.

Ji, J., et al., 2012. Proteolysis in microfluidic droplets: an approach to interface protein separation and peptide mass spectrometry. Lab Chip 12 (15), 2625−2629.

Jin, S., Anderson, G.J., Kenndy, R.T., 2013. Western blotting using microchip electrophoresis interfaced to a protein capture membrane. Anal. Chem. 85 (12), 6073−6079.

Jin, S., Kennedy, R.T., 2015. New developments in Western blot technology. Chin. Chem. Lett. 26 (4), 416−418.

Kalra, H., et al., 2013. Comparative proteomics evaluation of plasma exosome isolation techniques and assessment of the stability of exosomes in normal human blood plasma. Proteomics 13 (22), 3354−3364.

Karimi, A., Yazdi, S., Ardekani, A., 2013. Hydrodynamic mechanisms of cell and particle trapping in microfluidics. Biomicrofluidics 7 (2), 021501.

Karns, K., Herr, A.E., 2011. Human tear protein analysis enabled by an alkaline microfluidic homogeneous immunoassay. Anal. Chem. 83 (21), 8115−8122.

Khandurina, J., Chován, T., Guttman, A., 2002. Micropreparative fraction collection in microfluidic devices. Anal. Chem. 74 (7), 1737−1740.

Kim, D., et al., 2012. Electrostatic protein immobilization using charged polyacrylamide gels and cationic detergent microfluidic Western blotting. Anal. Chem. 84 (5), 2533−2540.

Kim, H., et al., 2013. A microfluidic DNA library preparation platform for next-generation sequencing. PLoS One 8 (7), e68988.

Kim, J., Johnson, M., Hill, P., Gale, B.K., 2009. Microfluidic sample preparation: cell lysis and nucleic acid purification. Integr. Biol. 574−586.

Kim, J., et al., 2010. A PCR reactor with an integrated alumina membrane for nucleic acid isolation. Analyst 135 (9), 2408−2414.

Kim, T.-H., Park, J., Kim, C.-J., Cho, Y.-K., 2014. Fully integrated lab-on-a-disc for nucleic acid analysis of food-borne pathogens. Anal. Chem. 86 (8), 3841–3848.

Klein, A.M., et al., 2015. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell 161 (5), 1187–1201.

Klostranec, J.M., et al., 2007. Convergence of quantum dot barcodes with microfluidics and signal processing for multiplexed high-throughput infectious disease diagnostics. Nano Lett. 7 (9), 2812–2818.

Konry, T., et al., 2011. Ultrasensitive detection of low-abundance surface-marker protein using isothermal rolling circle amplification in a microfluidic nanoliter platform. Small 7 (3), 395–400.

Kühnemund, M., Witters, D., Nilsson, M., Lammertyn, J., 2014. Circle-to-circle amplification on a digital microfluidic chip for amplified single molecule detection. Lab Chip 14 (16), 2983–2992.

Kuroda, A., et al., 2014. Microfluidics-based in situ padlock/rolling circle amplification system for counting single DNA molecules in a cell. Anal. Sci. 30 (12), 1107–1112.

Küster, S.K., et al., 2013. Interfacing droplet microfluidics with matrix-assisted laser desorption/ionization mass spectrometry: label-free content analysis of single droplets. Anal. Chem. 85 (3), 1285–1289.

Kwasny, D., et al., 2014. A semi-closed device for chromosome spreading for cytogenetic analysis. Micromachines 5 (2), 158–170.

Kwasny, D., et al., 2012. Microtechnologies enable cytogenetics. In: Tirunilai, P. (Ed.), Recent Trends in Cytogenetic Studies-Methodologies and Applications. INTECH Open Access Publisher, Rijeka, pp. 20–46.

Lagally, E.T., Simpson, P.C., Mathies, R.A., 2000. Monolithic integrated microfluidic DNA amplification and capillary electrophoresis analysis system. Sens. Actuators B Chem. 63 (3), 138–146.

Lapierre, F., et al., 2011. High sensitive matrix-free mass spectrometry analysis of peptides using silicon nanowires-based digital microfluidic device. Lab Chip 11 (9), 1620–1628.

Lazar, I.M., Kabulski, J.L., 2013. Microfluidic LC device with orthogonal sample extraction for on-chip MALDI-MS detection. Lab Chip 13 (11), 2055–2065.

Lazar, I.M., Trisiripisal, P., Sarvaiya, H.A., 2006. Microfluidic liquid chromatography system for proteomic applications and biomarker screening. Anal. Chem. 78 (15), 5513–5524.

Lee, A., 2013. The third decade of microfluidics. Lab Chip 1660–1661.

Lee, J., Soper, S.A., Murray, K.K., 2009. Microfluidics with MALDI analysis for proteomics—a review. Anal. Chim. Acta 649 (2), 180–190.

Lee, W., Fan, X., 2012. Intracavity DNA melting analysis with optofluidic lasers. Anal. Chem. 84 (21), 9558–9563.

Liu, S., Shi, Y., Ja, W.W., Mathies, R.A., 1999. Optimization of high-speed DNA sequencing on microfabricated capillary electrophoresis channels. Anal. Chem. 71 (3), 566–573.

Li, Z., et al., 2010. Rapid and sensitive detection of protein biomarker using a portable fluorescence biosensor based on quantum dots and a lateral flow test strip. Anal. Chem. 82 (16), 7008–7014.

Lochovsky, C., Yasotharan, S., Günther, A., 2012. Bubbles no more: in-plane trapping and removal of bubbles in microfluidic devices. Lab Chip 12 (3), 595–601.

Macosko, E.Z., et al., 2015. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161 (5), 1202–1214.

Mahmoudian, L., et al., 2008. Rolling circle amplification and circle-to-circle amplification of a specific gene integrated with electrophoretic analysis on a single chip. Anal. Chem. 80 (7), 2483–2490.

Mani, V., et al., 2009. Ultrasensitive immunosensor for cancer biomarker proteins using gold nanoparticle film electrodes and multienzyme-particle amplification. ACS Nano 3 (3), 585–594.

Manz, A., Graber, N., Widmer, H.M., 1990. Miniaturized total chemical analysis systems: a novel concept for chemical sensing. Sens. Actuators B Chem. 244–248.

Mao, X., Huang, T.J., 2012. Microfluidic diagnostics for the developing world. Lab Chip 12 (8), 1412–1416.

Mariella Jr., R., 2008. Sample preparation: the weak link in microfluidics-based biodetection. Biomed. Microdevices.

Melin, J., Melin, J., Johansson, H., Söderberg, O., Nikolajeff, F., Landegren, U., Nilsson, M., Jarvius, J., 2005. Thermoplastic microfluidic platform for single-molecule detection, cell culture, and actuation. Anal. Chem. 77 (22), 7122–7130.

Mery, E., et al., 2008. A silicon microfluidic chip integrating an ordered micropillar array separation column and a nano-electrospray emitter for LC/MS analysis of peptides. Sens. Actuators B Chem. 134 (2), 438–446.

Metzker, M.L., 2010. Sequencing technologies—the next generation. Nat. Rev. Genet. 11 (1), 31–46.

Mezger, A., et al., 2014. Detection of rotavirus using padlock probes and rolling circle amplification. PLoS One 9 (11), e111874.

Mitchell, P., Feldman, J., Pollock, N.R., Klapperich, C.M., 2012. Microfluidic chip for molecular amplification of influenza A RNA in human respiratory specimens. PLoS One 7 (3), e33176.

Mitchell, P.S., et al., 2008. Circulating microRNAs as stable blood-based markers for cancer detection. Proc. Natl. Acad. Sci. U. S. A 105 (30), 10513–10518.

Neuzil, P., et al., 2006. Ultra fast miniaturized real-time PCR: 40 cycles in less than six minutes. Nucleic Acids Res. 34 (11), e77.

Ng, A.H., Uddayasankar, U., Wheeler, A.R., 2010. Immunoassays in microfluidic systems. Anal. Bioanal. Chem. 397 (3), 991–1007.

Northen, T.R., et al., 2007. Clathrate nanostructures for mass spectrometry. Nature 449 (7165), 1033–1036.

Oblath, E.A., Henley, W.H., Alarie, J.P., Ramsey, J.M., 2013. A microfluidic chip integrating DNA extraction and real-time PCR for the detection of bacteria in saliva. Lab Chip 13 (7), 1325–1332.

Oxford Nanopore, 2016. VolTRAX: Rapid, Programmable, Portable, Disposable Sample Processor [Online] Available at: https://publications.nanoporetech.com/2016/05/26/voltrax-rapid-programmable-portable-disposable-sample-processor/.

Paegel, B.M., Blazej, R.G., Mathies, R.A., 2003. Microfluidic devices for DNA sequencing: sample preparation and electrophoretic analysis. Curr. Opin. Biotechnol. 14 (1), 42–50.

Paegel, B.M., et al., 2002. High throughput DNA sequencing with a microfabricated 96-lane capillary array electrophoresis bioprocessor. Proc. Natl. Acad. Sci.U. S. A 99 (2), 574–579.

Paliwal, A., Vaissière, T., Herceg, Z., 2010. Quantitative detection of DNA methylation states in minute amounts of DNA from body fluids. Methods 52 (3), 242–247.

Park, B.H., Kim, Y.T., Jung, J.H., Seo, T.S., 2014. Integration of sample pretreatment, μPCR, and detection for a total genetic analysis microsystem. Microchim. Acta 181 (13–14), 1655–1668.

Pekin, D., et al., 2011. Quantitative and sensitive detection of rare mutations using droplet-based microfluidics. Lab Chip 11 (13), 2156–2166.

Pennisi, E., 2016. Pocket DNA sequencers make real-time diagnostics a reality. Science 351 (6275), 800–801.

Perez-Toralla, K., et al., 2015. FISH in chips: turning microfluidic fluorescence in situ hybridization into a quantitative and clinically reliable molecular diagnosis tool. Lab Chip 15 (3), 811–822.

Pješčić, I., Crews, N., 2012. Genotyping from saliva with a one-step microdevice. Lab Chip 12 (14), 2514–2519.

Pješčić, I., et al., 2011. Real-time damage monitoring of irradiated DNA. Integr. Biol. 3 (9), 937–947.

Pješčić, I., Tranter, C., Hindmarsh, P.L., Crews, N.D., 2010. Glass-composite prototyping for flow PCR with in situ DNA analysis. Biomed. Microdevices 12 (2), 333–343.

Prakash, R., et al., 2014. Multiplex, quantitative, reverse transcription PCR detection of influenza viruses using droplet microfluidic technology. Micromachines 6 (1), 63–79.

Priye, A., Ugaz, V.M., 2016. Convective PCR thermocycling with smartphone-based detection: a versatile platform for rapid, inexpensive, and robust mobile diagnostics. In: Lu, C., Verbridge, S.S. (Eds.), Microfluidic Methods for Molecular Biology. Springer International Publishing, s.l., pp. 55–69.

Priye, A., et al., 2016. Lab-on-a-drone: toward pinpoint deployment of smartphone-enabled nucleic acid-based diagnostics for mobile health care. Anal. Chem. 88 (9), 4651–4660.

Quick, J., et al., 2016. Real-time, portable genome sequencing for Ebola surveillance. Nature 530 (7589), 228–232.

Ro, K.W., Liu, J., Knapp, D.R., 2006. Plastic microchip liquid chromatography-matrix-assisted laser desorption/ionization mass spectrometry using monolithic columns. J. Chromatogr. A 1111 (1), 40–47.

Romanov, V., Davidoff, S.N., Miles, A.R., Grainger, D.W., Gale, B.K., Benjamin, B.D., 2014. A critical comparison of protein microarray fabrication technologies. Analyst 139 (6), 1303–1326.

Safavieh, M., et al., 2014. A simple cassette as point-of-care diagnostic device for naked-eye colorimetric bacteria detection. Analyst 139 (2), 482–487.

Sainiemi, L., et al., 2012. A microfabricated micropillar liquid chromatographic chip monolithically integrated with an electrospray ionization tip. Lab Chip 12 (2), 325–332.

Samuel, R., et al., 2016. 40 cycle PCR using human genomic DNA. In: Less than 1 Minute. Dublin, Ireland, Proceedings of MicroTAS 2016.

Sanvicens, N., Pastells, C., Pascual, N., Marco, M.-P., 2009. Nanoparticle-based biosensors for detection of pathogenic bacteria. TrAC Trends Anal. Chem. 1243–1252.

Sato, K., Tachihara, A., Renberg, B., Mawatari, K., Sato, K., Tanaka, Y., Jarvirus, J., Nilsson, M., Kitamori, T., 2010. Microbead-based rolling circle amplification in a microchip for sensitive DNA detection. Lab Chip 10 (10), 1262–1266.

Sayad, A.A., et al., 2016. A microfluidic lab-on-a-disc integrated loop mediated isothermal amplification for foodborne pathogen detection. Sens. Actuators B Chem. 227, 600–609.

Schoenitz, M., Grundemann, L., Augustin, W., Scholl, S., 2015. Fouling in microstructured devices: a review. Chem. Commun. 8213.

Schuler, F., et al., 2016. Digital droplet PCR on disk. Lab Chip 16 (1), 208–216.

Shamsi, M.H., Choi, K., Ng, A.H., Wheeler, A.R., 2014. A digital microfluidic electrochemical immunoassay. Lab Chip 14 (3), 547–554.

Shih, S.C., et al., 2012. Dried blood spot analysis by digital microfluidics coupled to nanoelectrospray ionization mass spectrometry. Anal. Chem. 84 (8), 3731–3738.

Song, Y., et al., 2015. Size-based cell sorting with a resistive pulse sensor and an electromagnetic pump in a microfluidic chip. Electrophoresis 36 (3), 398–404.

Srbek, J., Eickhoff, J., Effelsberg, U., Kraiczek, K., van de Goor, T., Coufal, P., 2007. Chip-based nano-LC-MS/MS identification of proteins in complex biological samples using a novel polymer microfluidic device. J. Sep. Sci. 30 (13), 2046–2052.

Stumpf, F., et al., 2016. LabDisk with complete reagent prestorage for sample-to-answer nucleic acid based detection of respiratory pathogens verified with influenza A H3N2 virus. Lab Chip 16 (1), 199–207.

Sundberg, S.O., et al., 2014. Quasi-digital PCR: enrichment and quantification of rare DNA variants. Biomed. Microdevices 16 (4), 639–644.

Sun, Y., et al., 2015. A lab-on-a-chip system with integrated sample preparation and loop-mediated isothermal amplification for rapid and quantitative detection of *Salmonella* spp. in food samples. Lab Chip 15 (8), 1898–1904.

Tanaka, Y., et al., 2011. Single-molecule DNA patterning and detection by padlock probing and rolling circle amplification in microchannels for analysis of small sample volumes. Anal. Chem. 83 (9), 3352–3357.

Tehranirokh, M., Kouzani, A.Z., Francis, P.S., Kanwar, J.R., 2013. Microfluidic devices for cell cultivation and proliferation. Biomicrofluidics 7 (5), 051502.

Teh, S.-Y., Lin, R., Hung, L.-H., Lee, A.P., 2008. Droplet microfluidics. Lab Chip 8 (2), 198–220.

Thompson, A.M., et al., 2014. Microfluidics for single-cell genetic analysis. Lab Chip 14 (7), 3135–3142.

Thuy, T.T., Inganäs, M., Ekstrand, G., Thorsén, G., 2010. Parallel sample preparation of proteins, from crude samples to crystals ready for MALDI-MS, in an integrated microfluidic system. J. Chromatogr. B 878 (28), 2803–2810.

Tian, H., Hühmer, A.F., Landers, J.P., 2000. Evaluation of silica resins for direct and efficient extraction of DNA from complex biological matrices in a miniaturized format. Anal. Biochem. 283 (2), 175–191.

Tröger, V., N., K., Gärtig, C., Kuhlmeier, D., 2015. Isothermal amplification and quantification of nucleic acids and its use in microsystems. J. Nanomedicine Nanotechnol. 6 (282), 2.

Unger, M.A., et al., 2000. Monolithic microfabricated valves and pumps by multilayer soft lithography. Science 113–116.

Vergauwe, N., et al., 2011. A versatile electrowetting-based digital microfluidic platform for quantitative homogeneous and heterogeneous bioassays. J. Micromechanics Microengineering 21 (5), 054026.

Volpatti, L.R., Yetisen, A.K., 2014. Commercialization of microfluidic devices. Trends Biotechnol. 32 (7), 347–350.

Weisenberger, D.J., et al., 2008. DNA methylation analysis by digital bisulfite genomic sequencing and digital MethyLight. Nucleic Acids Res. 36 (14), 4689–4698.

Wheeler, A.R., et al., 2005. Digital microfluidics with in-line sample purification for proteomics analyses with MALDI-MS. Anal. Chem. 77 (2), 534–540.

Whitesides, G.M., 2006. The origins and the future of microfluidics. Nature 368–373.

Woolley, A.T., Mathies, R.A., 1995. Ultra-high-speed DNA sequencing using capillary electrophoresis chips. Anal. Chem. 67 (20), 3676–3680.

Xie, J., et al., 2005. Microfluidic platform for liquid chromatography-tandem mass spectrometry analyses of complex peptide mixtures. Anal. Chem. 77 (21), 6947–6953.

Xue, G., Pang, H.-M., Yeung, E.S., 2001. On-line nanoliter cycle sequencing reaction with capillary zone electrophoresis purification for DNA sequencing. J. Chromatogr. A 914 (1), 245–256.

Xu, Z.-R., et al., 2010. A miniaturized spatial temperature gradient capillary electrophoresis system with radiative heating and automated sample introduction for DNA mutation detection. Electrophoresis 31 (18), 3137–3143.

Yetisen, A.K., Volpatti, L.R., 2014. Patent protection and licensing in microfluidics. Lab Chip 2217–2225.

Yin, H., et al., 2005. Microfluidic chip for peptide analysis with an integrated HPLC column, sample enrichment column, and nanoelectrospray tip. Anal. Chem. 77 (2), 527–533.

Zhang, H., et al., 2013. Detection of low-abundance KRAS mutations in colorectal cancer using microfluidic capillary electrophoresis-based restriction fragment length polymorphism method with optimized assay conditions. PLoS One 8 (1), e54510.

Zhang, Y., et al., 2009. DNA methylation analysis on a droplet-in-oil PCR array. Lab Chip 9 (8), 1059–1064.

Zhu, Z., et al., 2012. Highly sensitive and quantitative detection of rare pathogens through agarose droplet microfluidic emulsion PCR at the single-cell level. Lab Chip 12 (20), 3907–3913.

Chapter 12

Protein Diagnostics by Proximity Ligation: Combining Multiple Recognition and DNA Amplification for Improved Protein Analyses

G. Cane, K.-J. Leuchowius, O. Söderberg, M. Kamali-Moghaddam, M. Jarvius, I. Helbing, K. Pardali, B. Koos, T. Ebai and U. Landegren
Uppsala University, Uppsala, Sweden

12.1 INTRODUCTION

Having the total human genome sequence information in our hands, and no longer only represented in our cell nuclei, we now face exciting opportunities to analyze all of the protein products encoded in our genomes. The rapidly improving ability to investigate large numbers of proteins with excellent precision in extensive series of patient samples at moderate cost, first in research and then clinically, will greatly affect health care through expanded opportunities for molecular diagnostics. Whereas genetic diagnostics mainly uncovers the potential to develop a disease, protein diagnostics can elucidate ongoing physiologic states, including diseases, because proteins are the main gene products constituting the phenotype. Therefore, protein-level analyses allow early diagnosis, serving to reveal subclinical, still treatable diseases. Future high-performance protein assays stand to improve the functional understanding of physiologic and pathologic processes, and measurements of amounts of specific proteins and their modifications and interactions can provide new approaches to protein diagnostics. Protein assays can also assist in the selection of optimal treatments and monitoring of disease progression, thus offering important benefits in health care. Methods intended for point-of-care or even used in households to diagnose signs of diseases such as infections should be highly sensitive, robust, easy to implement and use, and affordable. However, currently available methods for protein detection do not fulfill these requirements adequately. In particular, highly parallel approaches will be needed both in research and clinically. High-throughput techniques for proteome analyses are required to process the information embodied in large-scale collections of human samples such as population-wide biobanks. In particular, consecutive samples collected from the same individuals at different times represent valuable resources for the identification of diagnostic protein markers, provided appropriate high-throughput protein detection methods become available (Jackson and Banks, 2010).

Proteome analyses are complicated by the enormous concentration ranges of protein classes in samples such as plasma, and to a lesser extent within cells and tissues, which places stringent demands on the ability to distinguish proteins that may differ by many orders of magnitude in concentrations. This selectivity problem is further compounded by the presence of structurally similar but functionally different variants of proteins that frequently arise via mechanisms such as differential splicing, processing, and posttranslational modifications. Moreover, alterations in protein function are manifested not only by modulation of the expression level and their covalent modifications but also by their rearrangement in complexes, which necessitates the characterization of interactions among sets of proteins in parallel with individual protein molecules.

Whereas mass spectrometry has emerged as a powerful research tool for mapping the protein composition in biological samples, assays based on reagents with affinity for specific proteins remain the most promising avenue for

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00012-2
Copyright © 2017 Elsevier Ltd. All rights reserved.

protein-based diagnostics. However, this approach is limited by the difficulty of generating affinity reagents that combine high specificity and sensitivity at a moderate cost. Therefore, several initiatives are under way aiming to raise comprehensive sets of affinity reagents against large subsets of human proteins, and ultimately all of them (Taussig et al., 2013). However, because of the risk of cross-reactivity with antigenic epitopes presented by unintentionally targeted proteins, it is unlikely that practically monospecific affinity reagents can be developed except for relatively abundantly expressed proteins, for which risks of cross-reactivity are much lower. Moreover, individual binders can recognize only epitopes that represent a small part of a protein, ignoring any concomitant and diagnostically relevant remote epitopes, emerging from protein—protein interactions, posttranslational modification, protein truncation, or fusion. Accordingly, assays that combine simultaneous binding to multiple epitopes on target molecules are needed for enhanced selectivity and reduced risks of cross-reactivity.

In this chapter, we will provide a brief description of current methods for protein analyses, as a background to a description of a general molecular strategy for protein analyses, the proximity ligation, extension, and hybridization chain reaction assays, and we discuss future developments of these techniques.

12.2 BINDING THE PROTEOME

Binding reagents are of central importance in most methods used to study proteins. It is not possible to construct reagents capable of binding specific proteins on the basis of knowledge of the sequence of amino acid in the proteins. This is in stark contrast to the situation for DNA hybridization probes, in which simple rules for base pairing can be employed to design specific probes. Therefore, protein binding reagents are typically obtained by processes of selection, either in vivo or in vitro, from large repertoires of molecules with constant and variable sequence elements. Depending on how reagents are generated and selected, they may preferentially recognize denatured linear peptides or three-dimensional determinants. As a consequence, the reagents may differ in their suitability for assays in which proteins are either in a native and fully processed configuration, such as tissue homogenate and serum, or chemically and/or heat-denatured, such as in immunohistochemistry or sodium dodecyl sulfate—polyacrylamide gel electrophoresis (SDS-PAGE) and Western blotting.

Polyclonal and monoclonal antibodies are the most commonly used sources of protein binders. The diversity of antibodies generated by rearrangements of multiple gene segments, combined with junctional diversity and somatic hypermutation and gene conversion, provides a wide repertoire of antibodies for the selection of high-affinity binders against virtually any target protein. Besides in vivo immunizations, antigen binding fragments of antibody can be selected from large libraries in, e.g., bacteriophages, or expressed together with their encoding nucleotide sequences by ribosome or RNA display. These same in vitro selection techniques have permitted the development of a growing list of scaffold proteins, with variable sequences that can form protein-binding clefts and surfaces. Repeat protein binders such as designed ankyrin repeat proteins (DARPins) and armadillo repeat proteins are examples of scaffold proteins used as affinity binders (Binz et al., 2005; Helma et al., 2015). In addition, DNA- or RNA-aptamers can be selected from nucleic acid libraries to bind proteins with high affinity (Tuerk and Gold, 1990; Bock et al., 1992). Moreover, a library of natural nonimmunoglobulin binders is readily available including ligands that can be used to detect receptors, and glycan-binding proteins predominantly employed for the characterization of carbohydrate modifications of proteins (Daly and Mcgrath, 2003; Song et al., 2011). A number of projects have been initiated to establish comprehensive repertoires of binding agents against large numbers of identified human proteins, and eventually all of them. Examples of such initiatives include the Human Proteome Atlas (http://www.proteinatlas.org), developing polyclonal affinity-purified antibodies against human proteins; the National Cancer Institute, generating monoclonal antibodies against proteins of special interest in malignancy; and the European Union project Affinomics (http://www.affinomics.org), involving a number of mainly European groups, developing protein binding reagents (Stoevesandt and Taussig, 2012). It is therefore reasonable to assume that reliable clonal and replenishable binding reagents that can be widely shared among laboratories will be increasingly available for protein detection in years to come. Accordingly, assays must now be established where such reagents can be applied for extensive analysis of the proteome.

12.3 CURRENT AFFINITY-BASED PROTEIN DETECTION ASSAYS

Early protein-detection assays depended on target binding by single antibodies, such as the radioimmunoassays developed in the late 1950s (Yalow and Berson, 1960). Sandwich immunoassays, first published less than 10 years later, improved detection selectivity, sensitivity, and convenience by depending on the recognition of the target by two antibodies, one immobilized on a solid support and responsible for the initial capture of the antigen and the second added subsequently in solution and detectable via a reporter enzyme (Wide et al., 1967; Engvall and Perlmann, 1971). In a different manner, SDS-PAGE and Western

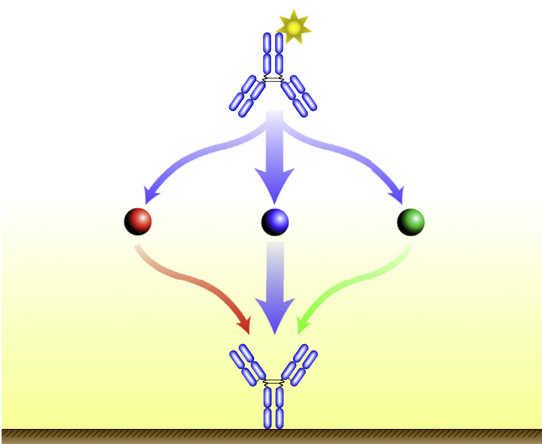
blotting enable the identification of proteins according to their electrophoretic mobility followed by immunoblotting with an antibody recognizing the protein target, thereby improving selectivity over assays exclusively dependent on recognition by a single-binder (Towbin et al., 1979). From early on, the approach to develop enzyme-linked immunosorbent assay (ELISA) and related methods further emphasized microtiter formats and automation for simultaneous analysis of large number of samples and protein analytes (Voller et al., 1974; Voller and O'Neill, 1971). The selectivity of detection in array-based assays can be improved by using a sandwich format with a pair of antibodies, but the rapidly increasing risks of cross-reactive binding among noncognate antibody pairs tend to limit such assays to a few tens of analytes (Fig. 12.1; Nielsen and Geierstanger, 2004). Highly sensitive sandwich immune assays have been developed to detect single protein molecules in which individual detection complexes are confined in femtoliter wells for digital detection (Rissin et al., 2010). In another approach to digital protein detection, antibody−target complexes are diluted in low cross-section capillaries to count single molecules that have been bound by fluorophore-labeled antibodies (Todd et al., 2007).

Immunohistochemistry is another widely used method in diagnostics that relies on the detection of proteins by primary or secondary antibodies covalently attached to detectable moieties such as fluorophores, rare earth metals, radioisotopes, or enzymes catalyzing a chromogenic reaction (Coons and Kaplan, 1950). Fluorescent dyes generate a signal proportional to the amount of bound antibodies, and thus increase the amount of fluorophores present per antibody can enhance the signal. However, for low-abundance proteins or in cases of high background fluorescence, signal amplification may be required to improve the signal-to-noise ratio. Reactions involving antibody-conjugated enzymes result in signal amplification when substrate molecules are converted into detectable products. The strength of the signal increases over time, potentially allowing the detection of lower concentrations of antigens, although generally also increasing nonspecific signals in parallel and thus limiting any improvement of signal over background.

Complementary to in situ detection methods, digital imaging and post-acquisition processing are also of interest by enabling quantification of the readout signal and demonstrating the distribution of target proteins in fixed cells and tissues (Stack et al., 2014).

12.4 PROXIMITY-DEPENDENT NUCLEIC ACID−BASED ASSAYS

The advantages of polymerase-based DNA amplification have been exploited in protein detection assays by conjugating single-stranded oligonucleotides to affinity probes. Thereby polymerase chain reaction (PCR) can be used to amplify detection signals exponentially from bound antibodies with attached DNA strands in immuno-PCR reactions (Sano et al., 1992). The mechanism lends itself to multiplex assays in which each antibody is conjugated to a unique DNA oligonucleotide that can be amplified by PCR and identified. The amount of each PCR product correlates to the amount of antibody bound to an antigen, and thus to the amount of target protein in a sample. For localized detection in, e.g., tissue sections, another method of DNA amplification, rolling circle amplification (RCA), can be used. A highly processive DNA polymerase such as the bacteriophage phi 29 DNA polymerase can generate a long



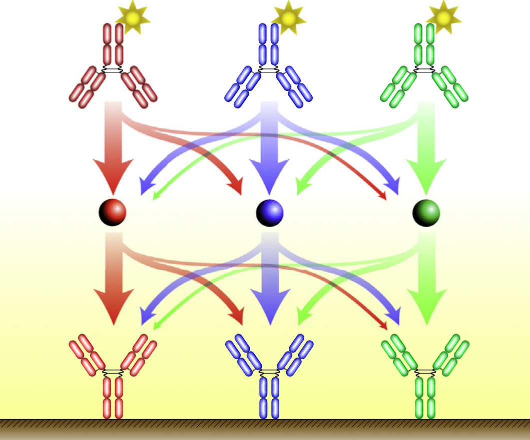


FIGURE 12.1 The risk for cross-reactive binding among noncognate antibody pairs increases when the number of proteins detected simultaneously in a sandwich immunoassay is increased. (A) When proteins are detected in singleplex, the risk of cross-reactive binding is low because target proteins must be recognized by both the capture antibody and the detection antibody for a positive identification. (B) In the case of multiplex detection, opportunities for cross-reactions increase rapidly. Any pair of capture and detection antibodies can result in detection signals, and whereas numbers of correct detection reactions increase linearly with increasing numbers of parallel assays, risks of cross-reactions grow roughly as the square of the numbers of reactions.

single-stranded concatemer of complements of a circular DNA strand used to template the reaction. The DNA molecule that forms can reach tens of micrometers in length but can be rolled up in a ball of DNA of submicrometer dimensions. The RCA technique has been used in immuno-RCA (Schweitzer et al., 2000; Wiltshire et al., 2000; Gusev et al., 2001), in which oligonucleotides are conjugated to antibodies and used to prime RCA reactions of subsequently added circular DNA strands. By hybridizing oligonucleotides labeled with fluorophores to the repeated sequences in the rolling circle products, the bundles of DNA can easily be visualized by fluorescence microscopy as brightly fluorescent spots, and the location of the binding event can be deduced.

A main consideration for all kinds of protein binding reagents is their selectivity for a target molecule. For positive identification of a protein, ideally the binder should interact with the correct protein with high affinity, and it should exhibit minimal cross-reactivity to other proteins. As discussed earlier, the selectivity of the detection reaction can be increased by adding a second selection criterion, such as the binding of a second antibody or by distinguishing target proteins according to their molecular weight or isoelectric charge, as performed in sandwich immune assays, Western blots, and after two-dimensional gel electrophoresis, respectively. In a related manner, proximity ligation reactions depend on proximal binding by two or more proximity ligation probes to increase the selectivity, whereas the sensitivity of the assays is improved by DNA amplification.

12.4.1 Proximity-Dependent Assays to Detect Proteins in Solution

In proximity ligation assays (PLA), single-stranded DNA oligonucleotides are conjugated to antibodies, and the conjugates are referred to as PLA probes. One antibody is coupled to an oligonucleotide via its 3′-end and another antibody is coupled to a second oligonucleotide via its 5′-end. When two PLA probes with different DNA-strands bind the same target protein or protein complex, a subsequently added connector oligonucleotide can hybridize to each free extremity of the conjugated DNA strands and guide their connection by enzymatic ligation. This creates a DNA molecule that can then be amplified and quantified, for instance by real-time PCR, whereas remaining non-ligated DNA strands cannot be amplified. By optimizing the design of the sequences to ensure that minimal ligation products form in the absence of target, subpicomolar levels of proteins can be detected by signal amplification through real-time PCR in a homogeneous assay with no need for washes or separations (Fredriksson et al., 2002; Gullberg et al., 2004; Fig. 12.2).

In a solid-phase variant of PLA, the target molecules to be detected are first captured by antibodies immobilized on a solid surface, as in sandwich ELISA, while buffer and excess components are removed by washes. Thereafter, a pair of PLA probes is added, unbound probes are removed by washes, and then the ligation reaction is performed followed by amplification of the connected DNA strand. The method provides further increased sensitivity of protein detection, and even single infectious units of microbial pathogens, virus, and bacteria can be demonstrated using this solid-phase format of PLA (Gustafsdottir et al., 2006). Generally, the solid-phase assay format has particular advantages for detecting rare molecules in dilute samples or in samples that contain components that can interfere with the detection reaction; the benefits of the homogeneous variant of PLA are the requirement for small amounts of sample and the simple and convenient assay procedure (Fredriksson et al., 2007). Both assay formats have been shown to permit multiplex analysis, because the DNA reactions can be used to ensure that only correct combinations result in detectable signals, whereas sandwich immune reactions, for instance, demonstrate increased cross-reactivity when more reactions are performed in parallel.

Other variants of PLA include the 3PLA and 4PLA. In 3PLA, the assays are designed to require target recognition by three PLA probes and two ligation events to form a complete PCR template (Schallmeiner et al., 2007). Similarly, the 4PLA requires the specific recognition of four different epitopes of the target by the PLA probes. The new molecule formed by three ligation events is then amplified and quantified by quantitative PCR (qPCR) (Tavoosidana et al., 2011). The requirement for simultaneous binding by several affinity binders decreases nonspecific background signal that may be caused by target-independent ligation or cross-reactivity for related target molecules, and hence even lower concentrations of the correct target protein can be detected over background signals. Furthermore, assays that depend on three binding events can be used to detect complex protein interactions or modifications involving three target epitopes (Schallmeiner et al., 2007). PLA has been successfully combined with array readout for parallel protein detection in a small-scale experiment, and the method is also suitable for identifying pairs of interacting proteins among a large set of proteins (Ericsson et al., 2008). The proximity ligation technique has also been adapted to examine interactions between proteins and specific DNA sequences (Gustafsdottir et al., 2007), which illustrates the versatility of this new detection mechanism. PLA has also been optimized for the simultaneous detection of 36 proteins using as little as 5 μL of plasma from patients with cardiovascular diseases as a proof of concept. Here, oligonucleotides carrying unique DNA tag sequences

FIGURE 12.2 (A) Schematic description of a homogeneous proximity ligation assay. (1) The target protein is recognized by the antibody parts of two different proximity ligation assay (PLA) probes, bringing their attached DNA strands in proximity during a first incubation. (2) Next, reactions are diluted and a connector oligonucleotide is added that can hybridize to both ends of DNA strands of pairs of PLA probes remaining in proximity by virtue of having bound the same target protein. Thus, the oligonucleotides can be joined by ligation, forming DNA reporter molecules for all detected proteins. (3) These DNA strands are then used for polymerase chain reaction (PCR) with real-time detection of the accumulating amplification products as a measure of the quantity of detected proteins. (B) An illustration of typical PLA-based real-time PCR results for quantification of tumor necrosis factor-α. *IFN*, interferon.

were conjugated to antibodies targeting the proteins of interest. After amplification via universal primer sequences, reporter strands for specific detected proteins were identified and quantified by qPCR or next-generation sequencing readout, the so-called ProteinSeq. The reagents and the midplex format readout have been selected with the intention of further increasing the number of measurements with the aim of reaching still higher multiplexing and throughput (Darmanis et al., 2011).

Assay formats based on mix and read procedure without sample processing and washing steps have advantages for molecular diagnostics. They are straightforward to execute and more reproducible and require less calibration and data processing for complex samples such as human plasma. The proximity extension assay (PEA) is a homogeneous assay for protein detection based on antibodies conjugated to single-stranded DNA oligonucleotides, as described previously for PLA. In PEA, short unique complementary

sequences located at the free 3′-end of each oligonucleotide hybridize to each other when brought in close proximity, allowing their subsequent polymerase-mediated extension. The oligonucleotides carry sequences targeted by PCR primers designed to be specific for each target protein enabling multiple analyte detection. An improvement in the method includes a preamplification step right after the extension via hybridization sequences for universal primers providing higher sensitivity of detection. Thereafter, a DNA polymerase generates amplicons that are subsequently distributed in a microfluidic device to amplify specific targets individually that are quantifiable by standard quantitative real-time PCR (Fig. 12.3; Lundberg et al., 2011; Assarsson et al., 2014). Optimization of the protocol introduced the hyperthermostable DNA polymerase Pwo in place of the Taq polymerase widely used in qPCR amplification, because the former also can serve to perform the initial extension reaction at room temperature without affecting the sensitivity of the assay, which makes the assay even more suitable for molecular diagnostics. So far, a 96-plex PEA immunoassay combined to microfluidic real-time PCR readout (Fluidigm) has been tested and validated with various complex samples routinely used in molecular diagnostics such as serum, plasma, dried blood spots, and tissue lysates (Assarsson et al., 2014). The assay can be carried out in a volume as little as 1 μL and has a femtomolar limit of detection and up to a 5-log dynamic range. It opens up a new horizon in molecular diagnostics with the possibility of detecting low abundant proteins in plasma samples (Lundberg et al., 2011). Furthermore, the assays

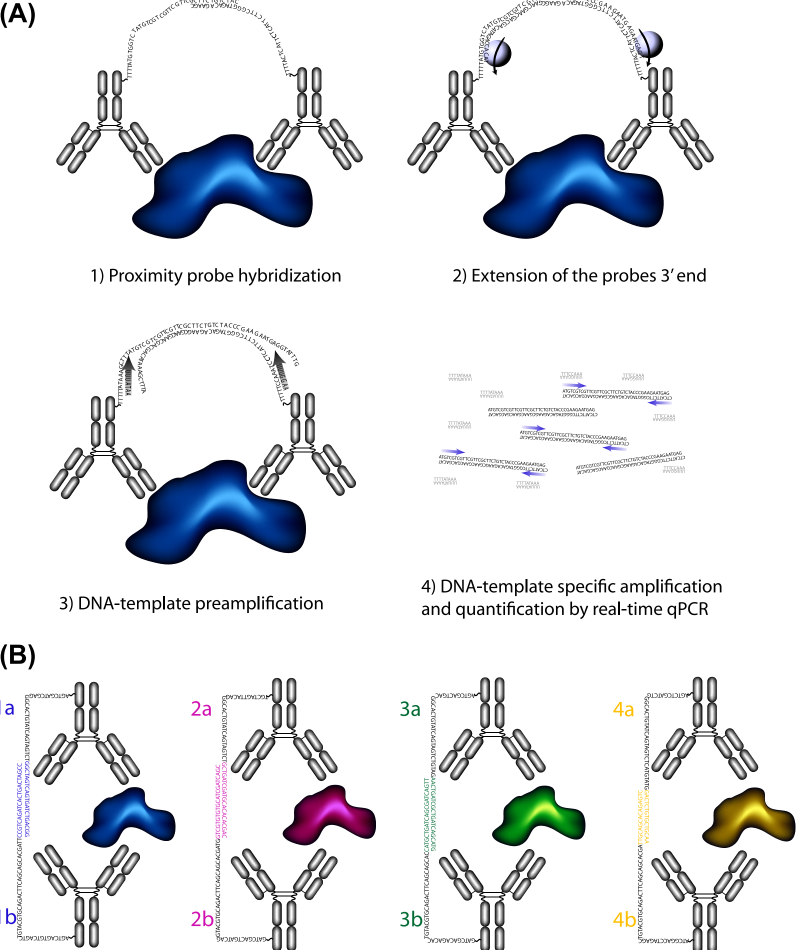


FIGURE 12.3 (A) Schematic representation of proximity extension assay (PEA). (1) Target proteins are recognized by pairs of PEA probes in a small reaction volume. (2) Because of the proximity of the probes binding the same target, the complementary 3′-ends of the antibody-conjugated DNA strands can anneal, allowing them to be extended by a DNA polymerase, and thus form a template for parallel preamplification (3). (4) Pairs of unique primers hybridize to the amplicons and the amplified products are detected and selectively quantified by real-time quantitative polymerase chain reaction (PCR) amplification with the Fluidigm® microfluidic device. (B) Schematic representation of multiplexed proximity extension assay (based on Olink Bioscience® scheme). Each unique pair of primers hybridizes only to their respective matched pair of probes, thus reducing the risk of cross-reactivity. The multiwell format allows multiplexed detection of almost 100 proteins and uses as little as a 1-μL sample. PEA is commercially available from Olink Bioscience® under the names of Proseek and Proseek Multiplex.

have been shown to reach sensitivities required for protein measurement in single cells (Darmanis et al., 2016). Since the development of the initial protocol, PEA has enabled the discovery of biomarkers for atherosclerosis, inflammatory bowel diseases, and colorectal neoplasia (Lind et al., 2015; Kalla et al., 2015; Thorsen et al., 2013). The assay has shown improved sensitivity of detection compared with currently used methods for detecting influenza An H1 viruses in nasopharyngeal swab samples (Van Wesenbeeck et al., 2013).

12.4.2 Proximity-Dependent Assays to Detect Proteins In Situ

Information about the subcellular location of protein and protein complexes and about their cell-to-cell variations within a tissue can provide important insights into the functional status of cells. PCR is not suitable for localized detection in cells or tissues, because PCR products tend to diffuse freely. Instead, by combining PLA with RCA, a prominent localized detection signal can be obtained at the site of specific recognition of individual target molecules, and with minimal background. For this purpose, samples are first incubated with pairs of PLA probes. If these probes bind in proximity, their attached DNA strands can serve as templates in a ligation reaction joining the ends of two outermost oligonucleotides that are included in the ligation reaction. This results in DNA circles that can next be replicated by RCA using one of the antibody-bound oligonucleotides as a primer. The localized amplification products are then visualized using fluorescence-labeled oligonucleotides that hybridize to the resulting RCA product (Soderberg et al., 2006; Fig. 12.4). The intensity of the brightly fluorescent spots, representing individual RCA products, reduces problems with background fluorescence. The individual fluorescence signals can be quantified digitally with the help of freely available software packages such as "BlobFinder," developed by Allalou and Wählby, and CellProfiler, developed at the Whitehead Institute for Biomedical Research and MIT's CSAIL (http://www.cb.uu.se/~amin/BlobFinder; http://www.cellprofiler.org/) for objective evaluation of the results.

The requirement for dual recognition by PLA probes greatly improves the selectivity of the method compared with immuno-RCA, because individual detection probes and other nucleic acids fail to result in circular DNA strands; as a consequence, no RCA products form unless dual recognition is achieved. The two PLA probes can be selected to bind to the same molecule for increased selectivity of detection, but they can also target two different proteins in a complex.

In the standard format of the assay, the distance between the bound PLA probes should be below tens of nanometers to give rise to detection signals. The antibodies add distance, so the targeted epitopes should be located within 20−40 nm, depending on whether primary or secondary PLA probes are used. Several different variants of in situ PLA have been developed (Koos et al., 2014) that extend the dynamic range (Clausson et al., 2011), reduce the size of the RCA products (Clausson et al., 2015), and facilitate parallel analysis of multiple proximity events (Leuchowius et al., 2013). Also, the use of DNA probes (Gomez et al., 2014; Weibrecht et al., 2012) and recombinant affinity reagents (Gu et al., 2013) as PLA probes have increased the numbers of possible targets for in situ PLA assays.

In situ PLA has been used successfully in hundreds of scientific papers to detect protein−protein interactions, targeting several signaling pathways, e.g., epidermal growth factor receptor/HER (Spears et al., 2012; Smith et al., 2015), transforming growth factor-β (Sundqvist et al., 2013; Zieba et al., 2012; Grannas et al., 2015), and Wnt (Figueiredo et al., 2013; Ribeiro et al., 2013; Huels et al., 2015) and for analysis of posttranslational modifications (PTMs) using pairs of antibodies in which one targets the core protein and the other a modified residue, e.g., phosphorylation (Jarvius et al., 2007) or glycosylation (Ricardo et al., 2015; Conze et al., 2010). Applications of in situ PLA for the detection of protein interactions have great potential for screening purposes in drug discovery and molecular diagnostics. The assay has the potential to be semiautomated and run in such a way so as to monitor for the formation or disruption of protein complexes, changes of subcellular localization, and simultaneous co-staining to facilitate the image analysis process (Leuchowius et al., 2010).

In situ PLA has also been used to study posttranslational modifications by using one antibody directed against platelet-derived growth factor receptor-β (PDGFR-β) and one antibody binding a phosphorylated tyrosine residue of the ligand stimulated receptor (Jarvius et al., 2007). In these experiments, samples were first incubated with a polyclonal rabbit antiserum against the receptor and a mouse monoclonal antibody directed against the phosphorylated residue. Next, two oligonucleotide-conjugated antibodies directed against rabbit and mouse immunoglobulin, respectively, were added to perform an in situ PLA. With this approach, standard oligonucleotide-conjugated secondary antibodies can be used for in situ PLA as long as the pair of primary antibodies is derived from two different animal species. This eliminates the need to construct antibody−oligonucleotide conjugates for all primary antibodies, thus saving time, cost, and effort.

Using in situ PLA to detect posttranslational modifications and, in particular, phosphorylations can also provide valuable information for molecular diagnostics and drug discovery and development. Detecting phosphorylation of a specific protein among other highly

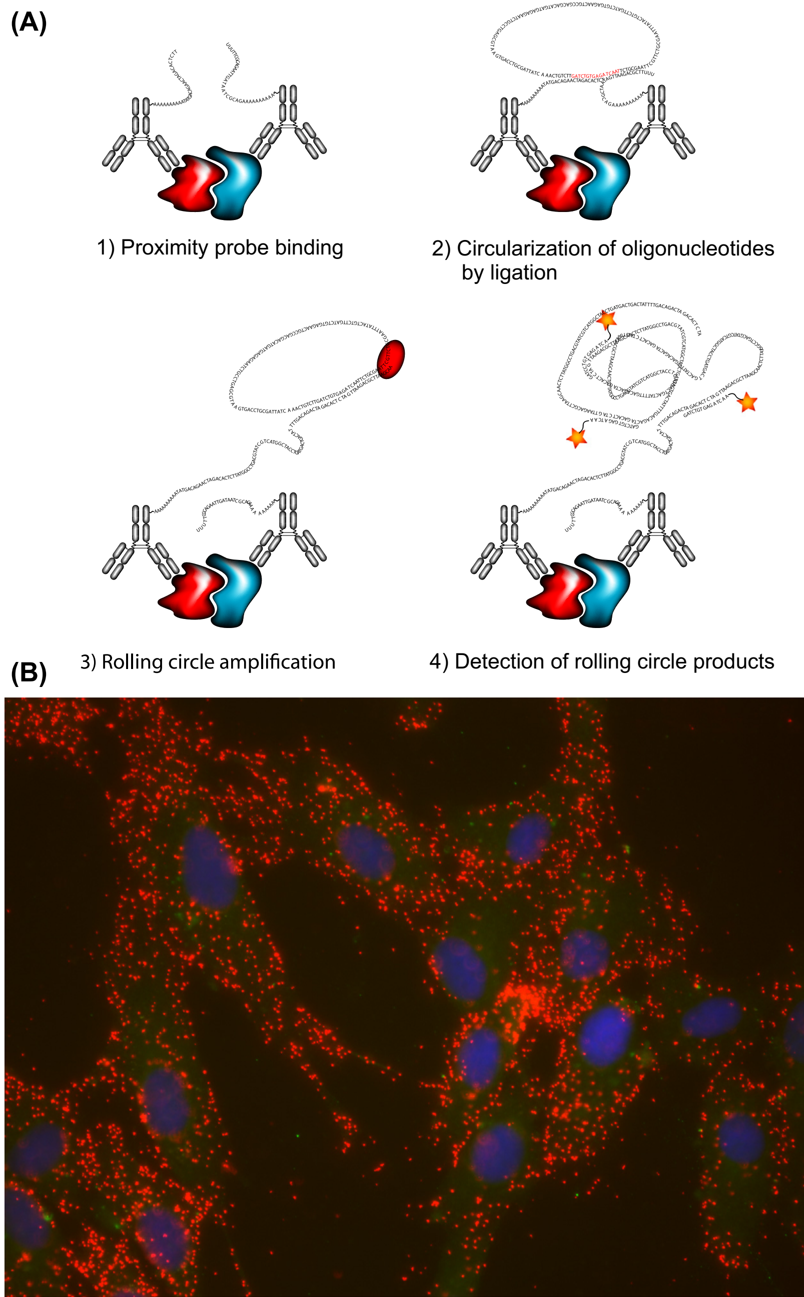


FIGURE 12.4 (A) Schematic presentation of in situ proximity ligation assay (PLA). Proximal binding by a pair of PLA probes (antibodies with attached DNA strands) to a target protein or pair of interacting proteins serves to template the formation of circular DNA molecules by ligation. The circular DNA molecules are then amplified by rolling circle amplification (RCA), primed by one of the PLA probes. The localized amplification reaction results in a concatemeric DNA product covalently attached to one of the proximity probes, and the RCA product can subsequently be visualized by hybridization of fluorescence-labeled oligonucleotides. (B) In situ PLA using secondary PLA probes detecting a pair of primary antibodies, directed against platelet-derived growth factor receptor-β and one of its phosphorylated tyrosine residues, respectively. Phosphorylated receptors are visualized as *red* (white in print versions) *dots* in cells that have been counterstained with antibodies directed against actin (green (gray in print versions)) and with the DNA binding Hoechst dye (blue (light gray in print versions)) to visualize the cytoplasm and the nucleus, respectively (Jarvius et al., 2007). PLA is commercially available from Olink Bioscience® under the name of Duolink.

homologous proteins in disease tissue in situ is a significant challenge. Examples of relevant proteins are members of evolutionary conserved family of receptors mediating immune-related diseases such as Toll-like receptors or downstream effector molecules such as component of the transcription factors of the nuclear factor-κB family proposed as biomarkers for various inflammatory diseases and cancers (Frans et al., 2014; Gambara et al., 2013). In

such cases, antibodies directed specifically against phosphorylated isoforms are generally not available because of the high degree of similarity of the site in the vicinity of the phosphorylated residue. Therefore, PLA remains the most accurate option to detect isoform-specific phosphorylation in situ.

In situ PLA has been applied to detect PTMs and messenger RNA (mRNA) molecules simultaneously in the same fixed cells. To achieve this, protocols were combined for protein detection via in situ PLA, and for mRNA detection via padlock probes. Upon complementary DNA (cDNA) synthesis templated by mRNA molecules of interest using locked nucleic acid–containing primers followed by RNase H digestion of the RNA strand, a single-stranded padlock probe complementary to the cDNA of interest was designed so that its 3′- and 5′-ends could hybridize to the cDNA strand in juxtaposition. Therefore, this oligonucleotide could be ligated to form a circular DNA molecule. An artificially introduced G/A mismatch between the padlock probe and the cDNA was used to cleave the cDNA at the hybridization site, permitting the priming of the RCA from the resulting free 3′-end. The RCA products resulting from the padlock probe used to detect mRNA and from the PLA probes detecting the PTM could be separately visualized using distinct fluorescence-labeled detection oligonucleotides (Weibrecht et al., 2011, 2013).

Cost-effectiveness and reimbursement by third-party payers are prerequisites in molecular diagnostics. Therefore, newly developed assays emphasize cost reductions through labor saving and reagents cost. In 2015, Koos and coworkers applied a previously developed enzyme-free DNA amplification method (Dirks and Pierce, 2004) to monitor protein–protein interaction and PTMs in situ (Koos et al., 2015a). The first step of the so-called proximity-dependent initiation of hybridization chain reaction (proxHCR) is the recognition of targeted epitopes by antibodies conjugated to oligonucleotides containing sequences complementary to each other but buried in a hairpin structure. Upon hybridization of a short initiator oligonucleotide to a sequence carried by one of the probes, the hairpin structure unfolds and exposes the complementary sequence. This triggers the invasion of the hairpin conjugated to the second antibody, provided that the pair of antibodies have bound to epitopes in close proximity. The subsequent chain reaction of events between pairs of fluorophore-labeled oligonucleotide hairpins generates a localized fluorescent product (Fig. 12.5). The method has been applied on cell lines as well as fresh-frozen and formalin-fixed, paraffin-embedded tissue material to detect phosphorylated proteins such as the PDGFR and the Syk protein and various protein–protein interactions such as E-cadherin/β-catenin known to serve as biomarker of cancer aggressiveness (Koos et al., 2015a). In a near future, proxHCR is expected to develop further because the method originally described has been optimized for multiplex analysis and real-time detection, which renders the procedure highly appealing for molecular diagnostics (Choi et al., 2011; Niu et al., 2010; Chemeris et al., 2008).

12.5 CONCLUSION AND FUTURE PERSPECTIVES

As discussed, opportunities for protein diagnostics are excellent. The development of new therapies will be increasingly accompanied by analysis of molecular markers, both nucleic acids and proteins, to evaluate clinical responses. Some of these assays will likely also follow the drugs into clinical routine use in the form of theranostic approaches. Molecular diagnostics can also be expected to become more broadly accessible through the development of simple new detection formats that vastly increase the scope for point-of-care diagnostics. The development of reagents and test architectures proceeds along several lines. Alternative recombinant affinity binders such as single chain fragments, nanobodies, DARPins, and affibodies can present important advantages because of their generally greater stability and easier production, as well as their clonal nature, ensuring reproducible production of binders. The combination of a wide variety of protein-binding reagents with information-carrying, amplifiable DNA strands can be a central element of many new diagnostic approaches. The mechanism provides valuable opportunities for increasing both the selectivity and sensitivity of detection. In contrast to more standard approaches, multiple proteins can be investigated in parallel without significantly increasing the risk of cross-reactivity by ensuring that only cognate pairs of proximity probes result in detection signals. Proximity-dependent methods lend themselves to assays where protein levels are measured and where their distribution is visualized in tissues, which illustrates the functional effects of signal transduction, developmental processes, and the effects of disease and of therapeutic intervention. A study demonstrated that the diagnostic value of protein measurements using PEA can be further enhanced by taking into account the genetic background of the sampled individuals, as well as information about age, sex, body mass index, etc. (Enroth et al., 2014). This study may herald a new trend of considering patient-specific ranges for protein concentrations, enhancing opportunities for identifying variations of proteins that are significantly associated with disease.

There are exciting opportunities for the further development of the proximity techniques. Measurement of proteins in solution is now performed simultaneously for 92 proteins and four controls in 1-μL sample aliquots using microfluidic readout. It is likely that higher

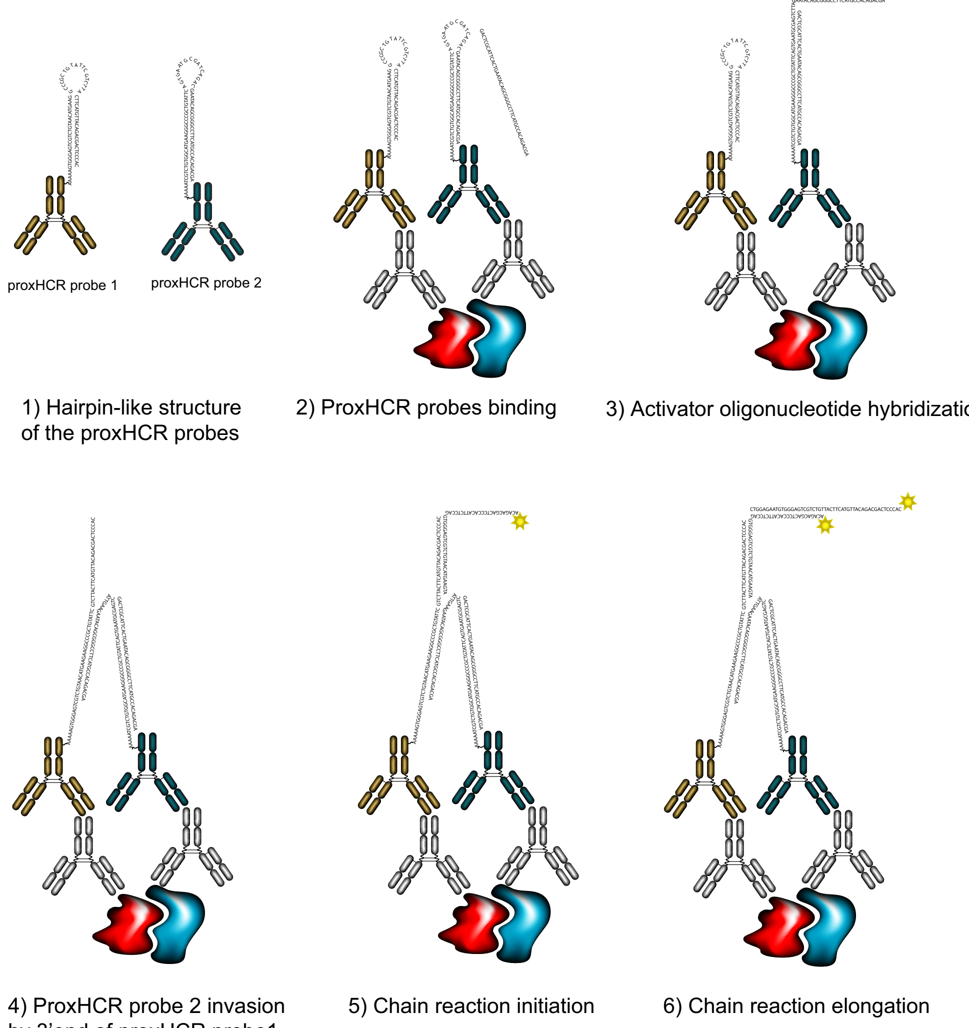


FIGURE 12.5 Schematic presentation of proximity-dependent initiation of hybridization chain (proxHCR). (1) The DNA strand of each of the two proxHCR probes is designed as a stem-loop structure named hairpin and is conjugated to secondary antibodies raised against different species, (2) thus recognizing corresponding host species of primary antibodies targeting two different epitopes on a same or different proteins. (3) The opening of the hairpin structure of proxHCR probe 1 strictly depends on the hybridization of an activator oligonucleotide to a sequence carries by the hairpin stem. (4) The 3′-end of proxHCR probe 1 thus freed can thereafter invade the hairpin loop of proxHCR probe 2 only if the probes have bound their target in close proximity. (5, 6) The invasion triggers the extrusion of the 3′-end of proxHCR probe 2 and initiates a chain of hybridization involving two species of fluorescent-labeled hairpin-shaped oligonucleotides complementary to each other. This reaction results in the local deposition of fluorescent oligonucleotides, resulting in bright focal signals at the site of binding by pairs of proxHCR probes.

multiplexing will be possible using other readout methods, and the assay is compatible with a wide range of readout formats. Examples of suitable readouts for multiplex reactions include microfluidic real-time amplification reactions (Fredriksson et al., 2007), microarray detection mechanisms (Ericsson et al., 2008), and techniques for digital recording of individual PLA or PEA reaction products using new parallel DNA sequencing techniques (Schuster, 2008) or by digital readout of single-protein molecules via RCA products (Ke et al., 2013). Also, the in situ forms of the assay can be adapted to permit higher multiplexing for parallel analysis of numerous proteins and their interactions and modifications in the same sample for a better understanding of cellular responses (Leuchowius et al., 2013). The diagnostic value of measuring multiple protein interactions and posttranslational protein modifications is just one of many existing opportunities that are currently emerging. By retaining architectural information combined with molecular profiling, the use of PLA and padlock probes in situ can provide information about the microenvironment in tissues, a level of detail that likely will be essential to the prediction of future pathology (Koos et al., 2015b).

ACKNOWLEDGMENT

The work in the Landegren lab is supported by grants from the Wallenberg Foundation, by the EU Integrated Project "ENLIGHT," Cancerfonden, and by the Research Councils of Sweden for Natural Science and Medicine. Ulf Landegren is a founder and share-holder of Olink Bioscience, commercializing proximity ligation and proximity extension assays.

REFERENCES

Assarsson, E., Lundberg, M., Holmquist, G., Bjorkesten, J., Thorsen, S.B., Ekman, D., Eriksson, A., Rennel Dickens, E., Ohlsson, S., Edfeldt, G., Andersson, A.C., Lindstedt, P., Stenvang, J., Gullberg, M., Fredriksson, S., 2014. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. PLoS One 9, e95192.

Binz, H.K., Amstutz, P., Pluckthun, A., 2005. Engineering novel binding proteins from nonimmunoglobulin domains. Nat. Biotechnol. 23, 1257−1268.

Bock, L.C., Griffin, L.C., Latham, J.A., Vermaas, E.H., Toole, J.J., 1992. Selection of single-stranded DNA molecules that bind and inhibit human thrombin. Nature 355, 564−566.

Chemeris, D.A., Nikonorov, Y.M., Vakhitov, V.A., 2008. Real-time hybridization chain reaction. Dokl. Biochem. Biophys. 419, 53−55.

Choi, J., Love, K.R., Gong, Y., Gierahn, T.M., Love, J.C., 2011. Immuno-hybridization chain reaction for enhancing detection of individual cytokine-secreting human peripheral mononuclear cells. Anal. Chem. 83, 6890−6895.

Clausson, C.M., Allalou, A., Weibrecht, I., Mahmoudi, S., Farnebo, M., Landegren, U., Wahlby, C., Soderberg, O., 2011. Increasing the dynamic range of in situ PLA. Nat. Methods 8, 892−893.

Clausson, C.M., Arngarden, L., Ishaq, O., Klaesson, A., Kuhnemund, M., Grannas, K., Koos, B., Qian, X., Ranefall, P., Krzywkowski, T., Brismar, H., Nilsson, M., Wahlby, C., Soderberg, O., 2015. Compaction of rolling circle amplification products increases signal integrity and signal-to-noise ratio. Sci. Rep. 5, 12317.

Conze, T., Carvalho, A.S., Landegren, U., Almeida, R., Reis, C.A., David, L., Soderberg, O., 2010. MUC_2 mucin is a major carrier of the cancer-associated sialyl-Tn antigen in intestinal metaplasia and gastric carcinomas. Glycobiology 20, 199−206.

Coons, A.H., Kaplan, M.H., 1950. Localization of antigen in tissue cells; improvements in a method for the detection of antigen by means of fluorescent antibody. J. Exp. Med. 91, 1−13.

Daly, C.J., Mcgrath, J.C., 2003. Fluorescent ligands, antibodies, and proteins for the study of receptors. Pharmacol. Ther. 100, 101−118.

Darmanis, S., Gallant, C.J., Marinescu, V.D., Niklasson, M., Segerman, A., Flamourakis, G., Fredriksson, S., Assarsson, E., Lundberg, M., Nelander, S., Westermark, B., Landegren, U., 2016. Simultaneous multiplexed measurement of RNA and proteins in single cells. Cell Rep 14, 380−389.

Darmanis, S., Nong, R.Y., Vanelid, J., Siegbahn, A., Ericsson, O., Fredriksson, S., Backlin, C., Gut, M., Heath, S., Gut, I.G., Wallentin, L., Gustafsson, M.G., Kamali-Moghaddam, M., Landegren, U., 2011. ProteinSeq: high-performance proteomic analyses by proximity ligation and next generation sequencing. PLoS One 6, e25583.

Dirks, R.M., Pierce, N.A., 2004. Triggered amplification by hybridization chain reaction. Proc. Natl. Acad. Sci. U.S.A. 101, 15275−15278.

Engvall, E., Perlmann, P., 1971. Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. Immunochemistry 8, 871−874.

Enroth, S., Johansson, A., Enroth, S.B., Gyllensten, U., 2014. Strong effects of genetic and lifestyle factors on biomarker variation and use of personalized cutoffs. Nat. Commun. 5, 4684.

Ericsson, O., Jarvius, J., Schallmeiner, E., Howell, M., Nong, R.Y., Reuter, H., Hahn, M., Stenberg, J., Nilsson, M., Landegren, U., 2008. A dual-tag microarray platform for high-performance nucleic acid and protein analyses. Nucleic Acids Res. 36, e45.

Figueiredo, J., Soderberg, O., Simoes-Correia, J., Grannas, K., Suriano, G., Seruca, R., 2013. The importance of E-cadherin binding partners to evaluate the pathogenicity of E-cadherin missense mutations associated to HDGC. Eur. J. Hum. Genet. 21, 301−309.

Frans, G., Meyts, I., Picard, C., Puel, A., Zhang, S.Y., Moens, L., Wuyts, G., Van Der Werff Ten Bosch, J., Casanova, J.L., Bossuyt, X., 2014. Addressing diagnostic challenges in primary immunodeficiencies: laboratory evaluation of Toll-like receptor- and NF-kappaB-mediated immune responses. Crit. Rev. Clin. Lab. Sci. 51, 112−123.

Fredriksson, S., Gullberg, M., Jarvius, J., Olsson, C., Pietras, K., Gustafsdottir, S.M., Ostman, A., Landegren, U., 2002. Protein detection using proximity-dependent DNA ligation assays. Nat. Biotechnol. 20, 473−477.

Fredriksson, S., Dixon, W., Ji, H., Koong, A.C., Mindrinos, M., Davis, R.W., 2007. Multiplexed protein detection by proximity ligation for cancer biomarker validation. Nat. Methods 4, 327−329.

Gambara, G., De Cesaris, P., De Nunzio, C., Ziparo, E., Tubaro, A., Filippini, A., Riccioli, A., 2013. Toll-like receptors in prostate infection and cancer between bench and bedside. J. Cell Mol. Med. 17, 713−722.

Gomez, A., Miller, N.S., Smolina, I., 2014. Visual detection of bacterial pathogens via PNA-based padlock probe assembly and isothermal amplification of DNAzymes. Anal. Chem. 86, 11992−11998.

Grannas, K., Arngarden, L., Lonn, P., Mazurkiewicz, M., Blokzijl, A., Zieba, A., Soderberg, O., 2015. Crosstalk between Hippo and TGFbeta: subcellular localization of YAP/TAZ/Smad complexes. J. Mol. Biol. 427 (21), 3407−3415.

Gu, G.J., Friedman, M., Jost, C., Johnsson, K., Kamali-Moghaddam, M., Pluckthun, A., Landegren, U., Soderberg, O., 2013. Protein tag-mediated conjugation of oligonucleotides to recombinant affinity binders for proximity ligation. N. Biotechnol. 30, 144−152.

Gullberg, M., Gustafsdottir, S.M., Schallmeiner, E., Jarvius, J., Bjarnegard, M., Betsholtz, C., Landegren, U., Fredriksson, S., 2004. Cytokine detection by antibody-based proximity ligation. Proc. Natl. Acad. Sci. U.S.A. 101, 8420−8424.

Gusev, Y., Sparkowski, J., Raghunathan, A., Ferguson Jr., H., Montano, J., Bogdan, N., Schweitzer, B., Wiltshire, S., Kingsmore, S.F., Maltzman, W., Wheeler, V., 2001. Rolling circle amplification: a new approach to increase sensitivity for immunohistochemistry and flow cytometry. Am. J. Pathol. 159, 63−69.

Gustafsdottir, S.M., Nordengrahn, A., Fredriksson, S., Wallgren, P., Rivera, E., Schallmeiner, E., Merza, M., Landegren, U., 2006. Detection of individual microbial pathogens by proximity ligation. Clin. Chem. 52, 1152−1160.

Gustafsdottir, S.M., Schlingemann, J., Rada-Iglesias, A., Schallmeiner, E., Kamali-Moghaddam, M., Wadelius, C., Landegren, U., 2007. In vitro analysis of DNA-protein interactions by proximity ligation. Proc. Natl. Acad. Sci. U.S.A. 104, 3067−3072.

Helma, J., Cardoso, M.C., Muyldermans, S., Leonhardt, H., 2015. Nanobodies and recombinant binders in cell biology. J. Cell Biol. 209, 633−644.

Huels, D.J., Ridgway, R.A., Radulescu, S., Leushacke, M., Campbell, A.D., Biswas, S., Leedham, S., Serra, S., Chetty, R., Moreaux, G., Parry, L., Matthews, J., Song, F., Hedley, A., Kalna, G., Ceteci, F., Reed, K.R., Meniel, V.S., Maguire, A., Doyle, B., Soderberg, O., Barker, N., Watson, A., Larue, L., Clarke, A.R., Sansom, O.J., 2015. E-cadherin can limit the transforming properties of activating beta-catenin mutations. EMBO J. 34 (18), 2321−2333.

Jackson, D.H., Banks, R.E., 2010. Banking of clinical samples for proteomic biomarker studies: a consideration of logistical issues with a focus on pre-analytical variation. Proteomics Clin. Appl. 4, 250−270.

Jarvius, M., Paulsson, J., Weibrecht, I., Leuchowius, K.J., Andersson, A.C., Wahlby, C., Gullberg, M., Botling, J., Sjoblom, T., Markova, B., Ostman, A., Landegren, U., Soderberg, O., 2007. In situ detection of phosphorylated platelet-derived growth factor receptor beta using a generalized proximity ligation method. Mol. Cell Proteomics 6, 1500−1509.

Kalla, R., Kennedy, N., Hjelm, F., Modig, E., Sundell, M., Andreassen, B., Bergemalm, D., Ricanek, P., Soderholm, J., Vatn, M., Halfvarson, J., Gullberg, M., Satsangi, J., Consortium, I. C, 2015. Proximity extension assay technology identifies novel serum biomarkers for predicting inflammatory bowel disease: IBD character consortium. J. Crohns Colitis 9, S146−S147.

Ke, R., Nong, R.Y., Fredriksson, S., Landegren, U., Nilsson, M., 2013. Improving precision of proximity ligation assay by amplified single molecule detection. PLoS One 8, e69813.

Koos, B., Andersson, L., Clausson, C.M., Grannas, K., Klaesson, A., Cane, G., Soderberg, O., 2014. Analysis of protein interactions in situ by proximity ligation assays. Curr. Top. Microbiol. Immunol. 377, 111−126.

Koos, B., Cane, G., Grannas, K., Lof, L., Arngarden, L., Heldin, J., Clausson, C.M., Klaesson, A., Hirvonen, M.K., De Oliveira, F.M., Talibov, V.O., Pham, N.T., Auer, M., Danielson, U.H., Haybaeck, J., Kamali-Moghaddam, M., Soderberg, O., 2015a. Proximity-dependent initiation of hybridization chain reaction. Nat. Commun. 6, 7294.

Koos, B., Kamali-Moghaddam, M., David, L., Sobrinho-Simoes, M., Dimberg, A., Nilsson, M., Wahlby, C., Soderberg, O., 2015b. Next-generation pathology−surveillance of tumor microecology. J. Mol. Biol. 427, 2013−2022.

Leuchowius, K.J., Jarvius, M., Wickstrom, M., Rickardson, L., Landegren, U., Larsson, R., Soderberg, O., Fryknas, M., Jarvius, J., 2010. High content screening for inhibitors of protein interactions and post-translational modifications in primary cells by proximity ligation. Mol. Cell Proteomics 9, 178−183.

Leuchowius, K.J., Clausson, C.M., Grannas, K., Erbilgin, Y., Botling, J., Zieba, A., Landegren, U., Soderberg, O., 2013. Parallel visualization of multiple protein complexes in individual cells in tumor tissue. Mol. Cell. Proteomics 12, 1563−1571.

Lind, L., Arnlov, J., Lindahl, B., Siegbahn, A., Sundstrom, J., Ingelsson, E., 2015. Use of a proximity extension assay proteomics chip to discover new biomarkers for human atherosclerosis. Atherosclerosis 242, 205−210.

Lundberg, M., Eriksson, A., Tran, B., Assarsson, E., Fredriksson, S., 2011. Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. Nucleic Acids Res. 39, e102.

Nielsen, U.B., Geierstanger, B.H., 2004. Multiplexed sandwich assays in microarray format. J. Immunol. Methods 290, 107−120.

Niu, S., Jiang, Y., Zhang, S., 2010. Fluorescence detection for DNA using hybridization chain reaction with enzyme-amplification. Chem. Commun. (Camb) 46, 3089−3091.

Ribeiro, A.S., Sousa, B., Carreto, L., Mendes, N., Nobre, A.R., Ricardo, S., Albergaria, A., Cameselle-Teijeiro, J.F., Gerhard, R., Soderberg, O., Seruca, R., Santos, M.A., Schmitt, F., Paredes, J., 2013. P-cadherin functional role is dependent on E-cadherin cellular context: a proof of concept using the breast cancer model. J. Pathol. 229, 705−718.

Ricardo, S., Marcos-Silva, L., Pereira, D., Pinto, R., Almeida, R., Soderberg, O., Mandel, U., Clausen, H., Felix, A., Lunet, N., David, L., 2015. Detection of glyco-mucin profiles improves specificity of MUC16 and MUC1 biomarkers in ovarian serous tumours. Mol. Oncol. 9, 503−512.

Rissin, D.M., Kan, C.W., Campbell, T.G., Howes, S.C., Fournier, D.R., Song, L., Piech, T., Patel, P.P., Chang, L., Rivnak, A.J., Ferrell, E.P., Randall, J.D., Provuncher, G.K., Walt, D.R., Duffy, D.C., 2010. Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations. Nat. Biotechnol. 28, 595−599.

Sano, T., Smith, C.L., Cantor, C.R., 1992. Immuno-PCR: very sensitive antigen detection by means of specific antibody-DNA conjugates. Science 258, 120−122.

Schallmeiner, E., Oksanen, E., Ericsson, O., Spangberg, L., Eriksson, S., Stenman, U.H., Pettersson, K., Landegren, U., 2007. Sensitive protein detection via triple-binder proximity ligation assays. Nat. Methods 4, 135−137.

Schuster, S.C., 2008. Next-generation sequencing transforms today's biology. Nat. Methods 5, 16−18.

Schweitzer, B., Wiltshire, S., Lambert, J., O'malley, S., Kukanskis, K., Zhu, Z., Kingsmore, S.F., Lizardi, P.M., Ward, D.C., 2000. Immunoassays with rolling circle DNA amplification: a versatile platform for ultrasensitive antigen detection. Proc. Natl. Acad. Sci. U.S.A. 97, 10113−10119.

Smith, M.A., Hall, R., Fisher, K., Haake, S.M., Khalil, F., Schabath, M.B., Vuaroqueaux, V., Fiebig, H.H., Altiok, S., Chen, Y.A., Haura, E.B., 2015. Annotation of human cancers with EGFR signaling-associated protein complexes using proximity ligation assays. Sci. Signal 8, ra4.

Soderberg, O., Gullberg, M., Jarvius, M., Ridderstrale, K., Leuchowius, K.J., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, L.G., Landegren, U., 2006. Direct observation of individual endogenous protein complexes in situ by proximity ligation. Nat. Methods 3, 995−1000.

Song, X., Lasanajak, Y., Xia, B., Heimburg-Molinaro, J., Rhea, J.M., Ju, H., Zhao, C., Molinaro, R.J., Cummings, R.D., Smith, D.F., 2011. Shotgun glycomics: a microarray strategy for functional glycomics. Nat. Methods 8, 85−90.

Spears, M., Taylor, K.J., Munro, A.F., Cunningham, C.A., Mallon, E.A., Twelves, C.J., Cameron, D.A., Thomas, J., Bartlett, J.M., 2012. In situ detection of HER2:HER2 and HER2:HER3 protein-protein interactions demonstrates prognostic significance in early breast cancer. Breast Cancer Res. Treat. 132, 463−470.

Stack, E.C., Wang, C., Roman, KA., Hoyt, C.C., 2014. Multiplexed immunohistochemistry, imaging, and quantitation: a review, with an assessment of Tyramide signal amplification, multispectral imaging and multiplex analysis. Methods 70, 46−58.

Stoevesandt, O., Taussig, M.J., 2012. European and international collaboration in affinity proteomics. N. Biotechnol. 29, 511−514.

Sundqvist, A., Zieba, A., Vasilaki, E., Herrera Hidalgo, C., Soderberg, O., Koinuma, D., Miyazono, K., Heldin, C.H., Landegren, U., Ten Dijke, P., Van Dam, H., 2013. Specific interactions between Smad proteins and AP-1 components determine TGFbeta-induced breast cancer cell invasion. Oncogene 32, 3606−3615.

Taussig, M.J., Schmidt, R., Cook, E.A., Stoevesandt, O., 2013. Development of proteome-wide binding reagents for research and diagnostics. Proteomics Clin. Appl. 7, 756−766.

Tavoosidana, G., Ronquist, G., Darmanis, S., Yan, J., Carlsson, L., Wu, D., Conze, T., Ek, P., Semjonow, A., Eltze, E., Larsson, A., Landegren, U.D., Kamali-Moghaddam, M., 2011. Multiple recognition assay reveals prostasomes as promising plasma biomarkers for prostate cancer. Proc. Natl. Acad. Sci. U.S.A. 108, 8809−8814.

Thorsen, S.B., Lundberg, M., Villablanca, A., Christensen, S.L., Belling, K.C., Nielsen, B.S., Knowles, M., Gee, N., Nielsen, H.J., Brunner, N., Christensen, I.J., Fredriksson, S., Stenvang, J., Assarsson, E., 2013. Detection of serological biomarkers by proximity extension assay for detection of colorectal neoplasias in symptomatic individuals. J. Transl. Med. 11, 253.

Todd, J., Freese, B., Lu, A., Held, D., Morey, J., Livingston, R., Goix, P., 2007. Ultrasensitive flow-based immunoassays using single-molecule counting. Clin. Chem. 53, 1990−1995.

Towbin, H., Staehelin, T., Gordon, J., 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. U.S.A. 76, 4350−4354.

Tuerk, C., Gold, L., 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249, 505−510.

Van Wesenbeeck, L., Meeuws, H., De Wolf, H., Stuyver, L., 2013. Evaluation of a proximity extension assay for the detection of H1 2009 pandemic influenza viruses. J. Virol. Methods 193, 77−84.

Voller, A., O'neill, P., 1971. Immunofluorescence method suitable for large-scale application to malaria. Bull. World Health Organ 45, 524−529.

Voller, A., Bidwell, D., Huldt, G., Engvall, E., 1974. A microplate method of enzyme-linked immunosorbent assay and its application to malaria. Bull. World Health Organ 51, 209−211.

Weibrecht, I., Grundberg, I., Nilsson, M., Soderberg, O., 2011. Simultaneous visualization of both signaling cascade activity and end-point gene expression in single cells. PLoS One 6, e20148.

Weibrecht, I., Gavrilovic, M., Lindbom, L., Landegren, U., Wahlby, C., Soderberg, O., 2012. Visualising individual sequence-specific protein-DNA interactions in situ. N. Biotechnol. 29, 589−598.

Weibrecht, I., Lundin, E., Kiflemariam, S., Mignardi, M., Grundberg, I., Larsson, C., Koos, B., Nilsson, M., Soderberg, O., 2013. In situ detection of individual mRNA molecules and protein complexes or post-translational modifications using padlock probes combined with the in situ proximity ligation assay. Nat. Protoc. 8, 355−372.

Wide, L., Bennich, H., Johansson, S.G., 1967. Diagnosis of allergy by an in-vitro test for allergen antibodies. Lancet 2, 1105−1107.

Wiltshire, S., O'malley, S., Lambert, J., Kukanskis, K., Edgar, D., Kingsmore, S.F., Schweitzer, B., 2000. Detection of multiple allergen-specific IgEs on microarrays by immunoassay with rolling circle amplification. Clin. Chem. 46, 1990−1993.

Yalow, R.S., Berson, S.A., 1960. Immunoassay of endogenous plasma insulin in man. J. Clin. Invest. 39, 1157−1175.

Zieba, A., Pardali, K., Soderberg, O., Lindbom, L., Nystrom, E., Moustakas, A., Heldin, C.H., Landegren, U., 2012. Intercellular variation in signaling through the TGF-beta pathway and its relation to cell density and cell cycle phase. Mol. Cell. Proteomics 11, M111013482.

Chapter 13

Application of Proteomics to Medical Diagnostics

K.M. Legg[1], H.E. McKiernan[1] and N. Reisdorph[2]

[1]Center for Forensic Science Research and Education, Willow Grove, PA, United States; [2]University of Colorado Denver, Aurora, CO, United States

13.1 INTRODUCTION

Proteomics can be broadly defined as the large-scale study of expressed proteins, which encompasses the identification and characterization of proteins and peptides. Subfields of proteomics include the analysis of posttranslational modifications (PTMs), quantitative proteomics, and clinical proteomics. The latter aims to identify and quantify molecular biomarkers of disease. Whereas this information can be used to understand disease mechanisms, a major goal of clinical proteomics is to develop highly specific and sensitive diagnostic laboratory assays. In the clinical setting, proteomics-based approaches offer the potential to transform patient care through the deployment of these assays, which ultimately can lead to personalized approaches for therapeutic intervention.

This chapter will review proteomics-based biomarker discovery strategies and examine how validation and verification of candidate molecules are conducted. Several examples of discovery-based and applied clinical proteomics applications will be discussed. In addition, challenges and shortcomings that exist in the adoption of proteomics-based approaches in the clinical laboratory will be evaluated. Finally, future approaches and technologies that aim to ease the transition of proteomics-based approaches into the clinical laboratory will be described.

13.2 CLINICAL IMPACT AND "PROTEOMICS" POTENTIAL

A multitude of "next-generation" methodologies are emerging as potential tools to provide the basis for rapid, reliable, noninvasive, and inexpensive clinical assays. Such assays are based on molecular markers and can be used to assess disease risks, diagnose and characterize diseases more rapidly, develop personalize therapies, and monitor the efficiency of treatments. These approaches can be based on several classes of molecules, including transcripts, metabolites, lipids, and proteins. For example, next-generation sequencing (NGS) (see Chapter 9) is currently used to diagnose infectious diseases (Yozwiak et al., 2012), subtype tumors based on genetic mutations (Honeyman et al., 2014), and characterize variants in the cystic fibrosis transmembrane regulator (*CFTR*) gene in cystic fibrosis screening (Baker et al., 2015). Small-molecule mass spectrometry (MS) has been used for well over a decade to identify metabolic disorders in newborns (Bennett, 2014). More recently, ovarian masses are being screened for cancer risk using a multiprotein enzyme-linked immunosorbent assay from proteins discovered using a proteomics approach (Rein et al., 2011). Finally, a multiplex RNA panel is currently used to assess the risk of prostate cancer recurrence after prostatectomy (Den et al., 2015).

Clearly, various "−omics" approaches offer the medical community powerful tools with which to improve human health. Although assays based on small molecules, genes, and transcripts are appropriate in many cases, there are several key advantages to protein-based diagnostic tests. Proteins can be described as the functional units of the cell; as such, they can be used to provide both direct and dynamic indicators of disease state. The entire human proteome encompasses a diverse and broad set of molecules; this includes sequence variants, pro- and mature proteins, and a plethora of PTMs (Anderson et al., 2004b). Protein abundances differ according to organ, cell, and subcellular location; they are affected differentially by a variety of

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00013-4
Copyright © 2017 Elsevier Ltd. All rights reserved.

conditions including circadian rhythms, age, health status, and diet. Proteins have been analyzed in the context of many human conditions and have even been isolated from ancient biological material (Cappellini et al., 2011; Lindgren et al., 2011). All told, proteins can be used to determine mechanisms, elucidate pathways, and indicate disease status. For the latter, several clinical assays exist that are based on proteins; these include markers of inflammation (C-reactive protein, CRP), cardiovascular risk (brain natriuretic peptide, BNP), and prostate-specific antigen (PSA).

With the potential of protein-based diagnostic assays to improve physician care, proteomics-based discovery projects have exploded in recent years. Global, discovery-based proteomics studies can routinely identify several thousand proteins in a single experiment while flagging candidates that may be associated with disease (Nagaraj et al., 2011; Beck et al., 2011). Despite efforts spent on clinical proteomics discovery studies, a few novel protein biomarkers have been translated into clinical assays (Polanski and Anderson, 2006; Li and Chan, 2014). Clearly, a divide exists among the discovery of biomarker candidates, validation of candidates as bona fide markers, and implementation of proteomics-based diagnostic tests (Fig. 13.1). Interestingly, tests that have been developed often rely on conventional single-plex assays; these have been the reference standard for qualitative and quantitative detection of proteins in the clinical laboratory for over 50 years (Tighe et al., 2015). Nonetheless, there is potential for MS-based methodologies not only to identify biomarker candidates but also to be used for validation and as clinical assays.

A typical clinical proteomics pipeline follows an MS-based workflow that includes candidate discovery, validation in a larger cohort, and finally implementation as a clinical assay. The specific MS platform can vary from one phase to the next. Beginning with the discovery phase, proteins from a small set of biological samples are compared; the result is a panel of differentially expressed peptides and proteins. Within these panels, which may include several hundred candidates, it is common to observe a high false discovery rate. This is primarily because the small sample size is unable to account fully for inter- and intraindividual sample variability (Frantzi et al., 2014). After candidate identification, a quantitative verification phase is employed to confirm biological significance across a much larger sample population to assess biomarker specificity (Cancer, 2009). This acts as a critical "proof of association" among biomarker detection, expression, and disease association. Moving forward with the most promising candidates, the final diagnostic platform must be developed and analytically validated in accordance with generally accepted bioanalytical validation guidelines as well as Food and Drug Administration (FDA) guidelines (Services, 2015; Convention, 2015). This workflow is described in more detail subsequently.

13.3 STRATEGIES FOR MASS SPECTROMETRY–BASED PROTEOMICS: DISCOVERY AND VERIFICATION

13.3.1 Basic Workflow

Advances in instrumentation, computational abilities, and software have enabled the identification of several thousand proteins in a single discovery experiment; many of these are within physiologically relevant detection limits (Beck et al., 2011; Nagaraj et al., 2011; Percy et al., 2013). The discovery-to-verification workflow (Fig. 13.2) entails three major steps: sample preparation, including enzymatic digestion; data acquisition using mass spectrometers; and informatics, including protein identification, quantitation, and statistical analysis. At the discovery stage, sample sources can include biological fluids such as plasma, saliva, bronchoalveolar lavage fluid, sputum, and proximal fluids; tissues, cells lines, and animal models are also used at this early phase (Wattiez and Falmagne, 2005; Magi et al.,

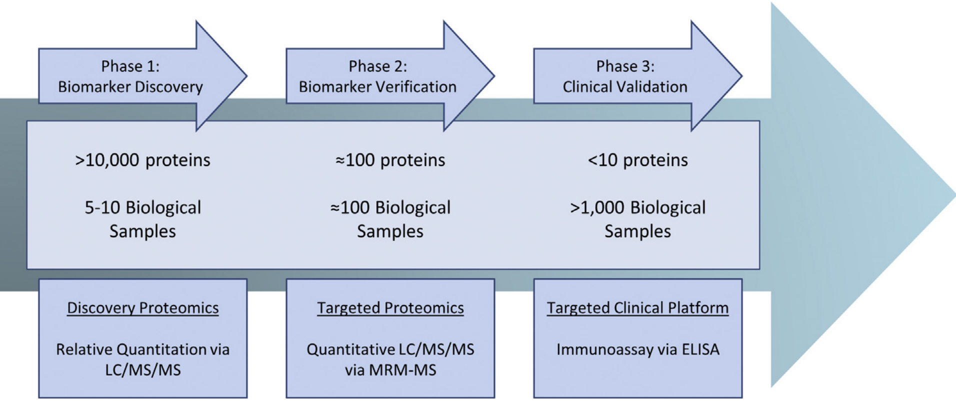


FIGURE 13.1 Three-stage biomarker pipeline from candidate identification to the development of a clinical assay. Pipeline includes biomarker candidate discovery, biomarker candidate verification, and finally the development of an immunoassay-based diagnostic test. *ELISA*, enzyme-linked immunosorbent assay; *LC/MS/MS*, liquid chromatography tandem mass spectrometry; *MRM-MS*, multiple-reaction monitoring–mass spectrometry.

2006; Gray et al., 2008; Nicholas et al., 2006). After candidate selection, empirically or literature-derived biomarkers must be verified using a targeted, hypothesis-driven, analytical workflow in a reference standard biological sample. The main verification technique currently used is multiple-reaction monitoring (MRM)-MS coupled with custom-synthesized, isotopically labeled internal peptide standards. Using this approach, several dozen protein candidates can be quantified in a single multiplex method (Percy et al., 2012a). The most promising candidate biomarkers then proceed to clinical qualification where the final ligand binding–based assay is developed and tested against several thousand biological samples to ensure diagnostic impact in a clinical setting.

13.3.2 Sample Preparation

Sample preparation protocols will vary depending on the sample source and the goals of the study. One primary challenge of proteomics-based research is the enormous number of different proteins that are present within biological fluids. Human plasma, for example, contains thousands of different proteins and spans a concentration range across nine orders of magnitude (Anderson, 2005). Albumin and immunoglobulin G make up roughly 50% of the total plasma protein content, and transferrin, fibrinogen, haptoglobin, and alpha-1 antitrypsin another 25% (Putnam, 1975). As a consequence, most highly abundant proteins, which may not be diagnostically relevant, can easily mask the detection of less abundant proteins through ion suppression during mass spectral analysis. To resolve this, highly abundant protein depletion using immunoaffinity columns is a standby technique for allowing the detection of lower-abundant biomarkers (Patel et al., 2012; Hakimi et al., 2014). Additional fractionation approaches, which may be used independently or in tandem with other techniques, exploit the physiochemical properties of the peptide or protein target. The three predominant techniques are molecular weight fractionation via sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE)

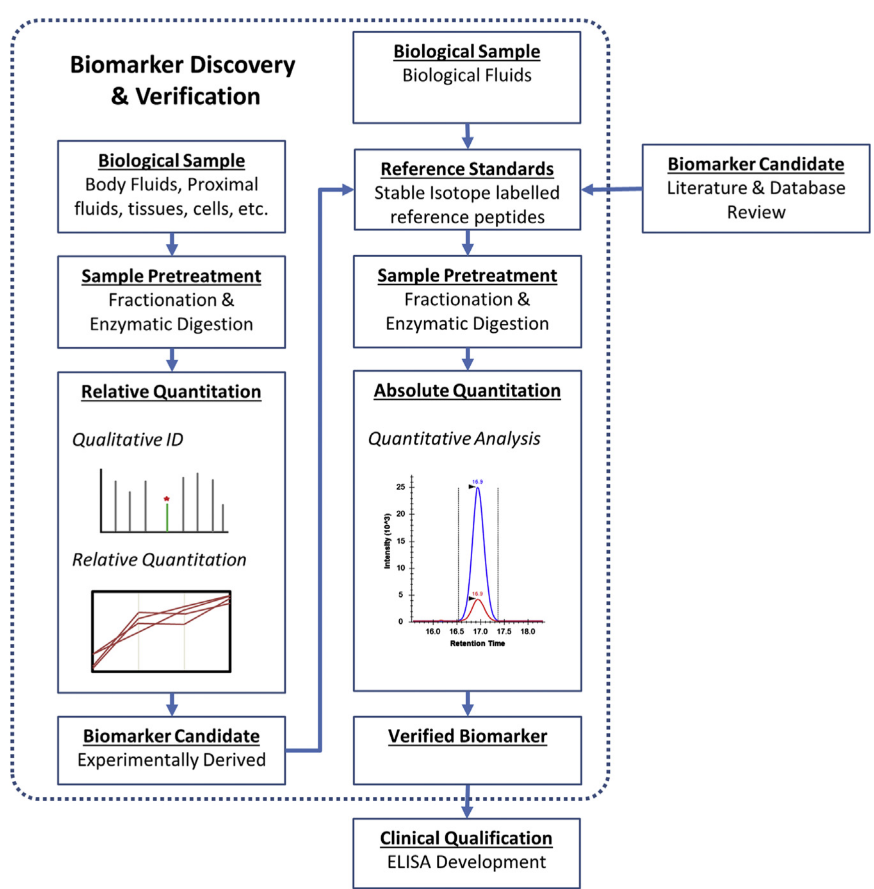


FIGURE 13.2 Biomarker discovery and verification approaches. Candidate discovery is carried out using biological specimens that are fractionated, enzymatically digested, and analyzed using untargeted, semiquantitative mass spectrometry to identify differentially expressed biomarker candidates. Conversely, verification workflows monitor specific biomarker candidates using targeted, quantitative mass spectrometry in conjunction with isotopically labeled internal standards. *ELISA*, enzyme-linked immunosorbent assay.

(Hardt et al., 2005), ion exchange via strong cation exchange (SCX) (Peng et al., 2003; Chaerkady et al., 2008), and pH fractionation with isoelectric focusing (IEF) (Cargile et al., 2005; Seshi et al., 2011). After fractionation, tryptic peptides are analyzed by one of three semi-quantitative workflows: comparative gel-based quantitation, label-based quantitation, or label-free quantitation, the results of which will flag one or more differentially expressed protein biomarkers.

13.3.3 Instrumentation

MS is the principal tool enabling all aspects of modern proteomics-based analysis. After sample fractionation and digestion, proteins are digested with a protease such as trypsin. Mass spectrometers, as the name infers, measure the mass of molecules. More specifically, an MS system measures the mass-to-charge ratio (m/z) of a charged molecule (Herbert and Johnstone, 2002; Glish and Vachet, 2003). In protein-based MS, this allows for the viability of qualitative and quantitative measurements of charged proteins or peptides (Domon and Aebersold, 2006). Protein MS commonly comes in two varieties (Fig. 13.3), based purely on the ionization source that generates charged molecules: matrix-assisted laser desorption/ionization time-of-flight MS (MALDI-TOF) or nanoflow-electrospray tandem MS (nLC/MS/MS).

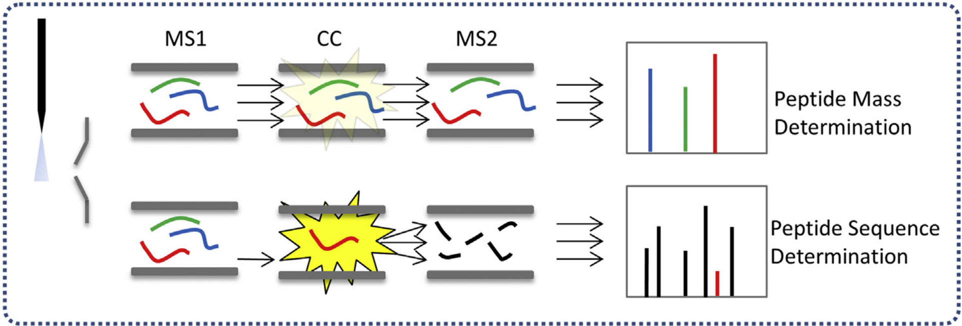
Using MALDI-TOF for peptide analysis, peptides are immobilized onto a metal plate combined with a matrix material consisting of a weak organic acid (e.g., α-cyano-4-hydroxycinnamic acid, sinapinic acid, or 2,5-dihydroxybenzoic acid). Ionization occurs when a nitrogen laser at a wavelength of 337 nm is fired into the dried peptide−matrix sample. This causes a portion of the sample−matrix to vaporize; the acidic matrix plume will protonate the peptides in the sample (Hillenkamp et al., 1986). Once ionized, positively charged peptides are uniformly accelerated in the TOF mass analyzer, where individual ions travel at a rate in proportion to their individual masses, with smaller masses traveling the fastest. The time it takes the peptides to reach the detector is recorded and mass spectrum is generated for the charged species (Lewis et al., 2000).

Coupling the mass spectrometer mass analyzer via a nanoflow-electrospray source is another popular technique. Here, a liquid-based peptide mixture is fractionated online and directly ionized from the liquid phase before MS analysis. As with any chromatographic technique, nLC/MS/MS uses a mobile phase as well as a stationary phase. A peptide digest is dissolved in the mobile phase and is then passed through a stationary phase. Every compound in a mixture will have a higher or lower affinity for the mobile and stationary phases, thus allowing complex mixtures to be separated. Common for proteomics studies, nanoflow chromatography is performed to maximize the detection of low-abundant targets from small sample quantities. This technique employs microdiameter columns (0.075 mm versus ≥2.1 mm inner diameter) as well as low flow rates (100 nL/min versus ≥400 μL/min) compared with traditional "high-flow" analysis. Because a mass spectrometer acts as a concentration-dependent analyzer, low flow rates increase the analyte concentration enabling much higher ionization efficiency, thus increasing the detection limit of the analysis method (Fröhlich and Arnold, 2009). After chromatographic separation, an electrospray ionization source generates positively charge peptides directly from the liquid effluent. In this process, the acidic eluent (i.e., a mixture of water, acetonitrile, and formic acid) from the nanoflow-electrospray is forced through a fine capillary needle while being exposed to superheated nitrogen drying gas and high voltage. Under these conditions, positively charged microdroplets migrate toward the MS inlet. Rapid drying of these charged microdroplets combined with electrostatic fission leads to protonated gas-phase molecules that enter the mass spectrometer (Kebarle and Verkerk, 2009). Once ionized, a tandem mass spectrometer will elucidate intact peptide masses (MS) and generate sequence-specific fragmentation spectra (MS/MS). More specifically, in standard data-dependent acquisition mode, the instrument generates a survey scan of all eluting peptides (MS scan). Acquisition software then selects specific precursor ions from the MS scan and the instrument will filter for one specific charged peptide. The selected precursor ion is then fragmented in the collision cell and the product ions are further resolved before detection (MS/MS). The precursor isolation, fragmentation, separation, and detection process produces the MS/MS fragmentation spectra for each peptide, generating amino acid sequence information.

In comparing these two approaches, using a MALDI-TOF system, immobilized peptide sample and matrix are ionized and analyzed to detect individual peptide masses. Using nLC/MS/MS, peptides are first separated on a nanocolumn and ionized using electrospray ionization; their mass is detected and then a peptide is fragmented in a mass analyzer, generating amino acid sequence information. MALDI-TOF is generally considered a high-throughput platform but it is limited to simple sample types, whereas nLC/MS/MS is generally more sensitive and amenable to the analysis of complex mixtures due to the online chromatographic separation.

13.3.4 Quantitative Mass Spectrometry

Quantitative MS-based experiments can be performed using cells, tissue biopsies, or biological fluids. Quantitation that relies on internal standards and results in a concentration-based value (i.e., pg/mL) is termed "absolute



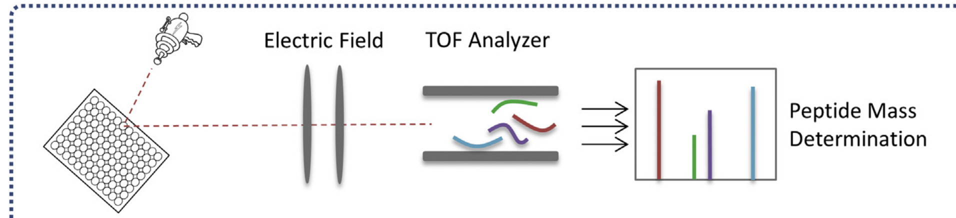


FIGURE 13.3 Protein mass spectrometry (MS) platforms including nanoflow-electrospray tandem MS (nLC/MS/MS) (*top*) and matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) MS (*bottom*). Using nLC/MS/MS, peptides are fractionated via inline reverse-phase chromatography before ionizing the liquid-phase eluent. Once ionized, the instrument will record intact (MS) peptide mass in conjunction with a full-fragmentation spectrum (MS/MS). MALDI-TOF systems ionize immobilized peptides suspended in a matrix. Once ionized by a nitrogen laser, charged peptides are analyzed by the mass spectrometer to elucidate intact peptide masses (MS). *CC*, collision cell.

quantitation." Quantitative methods that compare the relative amount of protein present in one sample versus another and result in a fold-change difference are termed "relative quantitation." Discovery-based workflows employ relative quantitation; this can be further divided into labeling or nonlabeling strategies. After discovery, biomarker validation is generally performed using absolute quantification on a triple quadrupole mass spectrometer. Fig. 13.4 is a graphical overview of each approach, which will be discussed independently.

13.3.4.1 Relative Quantitation Strategies for Discovery

Gel Electrophoresis: Historically, proteome analysis was performed using multidimensional gel electrophoresis to identify differentially expressed protein "spots" before identification by tandem MS. Using this approach, intact proteins are resolved in sequence using IEF followed by SDS-PAGE. Resolved proteins are stained (*e.g.*, Coomassie brilliant blue), analyzed for differential expression among several samples, excised, and analyzed by MS. Improvements to this traditional approach have used multiple fluorescent dyes (cyanine dyes Cy2, Cy3, and Cy5) to analyze multiple samples on a single gel. Called 2D difference gel electrophoresis (2D-GE), up to three samples are fluorescently labeled, mixed, and fractionated, and distinct images for each sample are generated from a single separation. Measuring the signal intensity for individual protein spots can result in the identification of statistically relevant, differentially expressed proteins between analyzed samples (Van den Bergh and Arckens, 2004). This approach is known to be more reliable because samples are resolved on a single gel, which eliminates variability inherent in 2D-GE studies. However, variability among gels is still a problem that makes it difficult to distinguish between actual protein variability versus gel-to-gel experimental variation (Alban et al., 2003). In addition, there limitations impede the accurate analysis of proteins with very high or low pI values, proteins that are insoluble in fractionation buffers, and large hydrophobic or small hydrophilic proteins (Liu et al., 2002; Schlautman et al., 2008).

In addition to 2D-GE, multidimensional fractionation by liquid chromatography offers another approach. Two examples, both using 2D high-performance liquid chromatography (2D-HPLC), are the ProteoSep™ system and Multidimensional Protein Identification Technology (mudPIT). Both have migrated from gel-based separations to precise and reproducible HPLC-based separations. Using the ProteoSep™ system, proteins are separated via inline chromatofocusing and reverse-phase chromatography. During the reverse-phase runs, a 214-nm UV detector monitors eluents as they are released from the column, allowing quantitative measurements for individual proteins while storing samples in liquid phase for rapid protein identification (Wang and Zhong, 2012). mudPIT employs 2D liquid chromatography separation technology based on SCX coupled with reverse-phase chromatography. What makes this methodology unique is that both reverse-phase and SCX resins are packaged together in a single microcapillary column with a direct

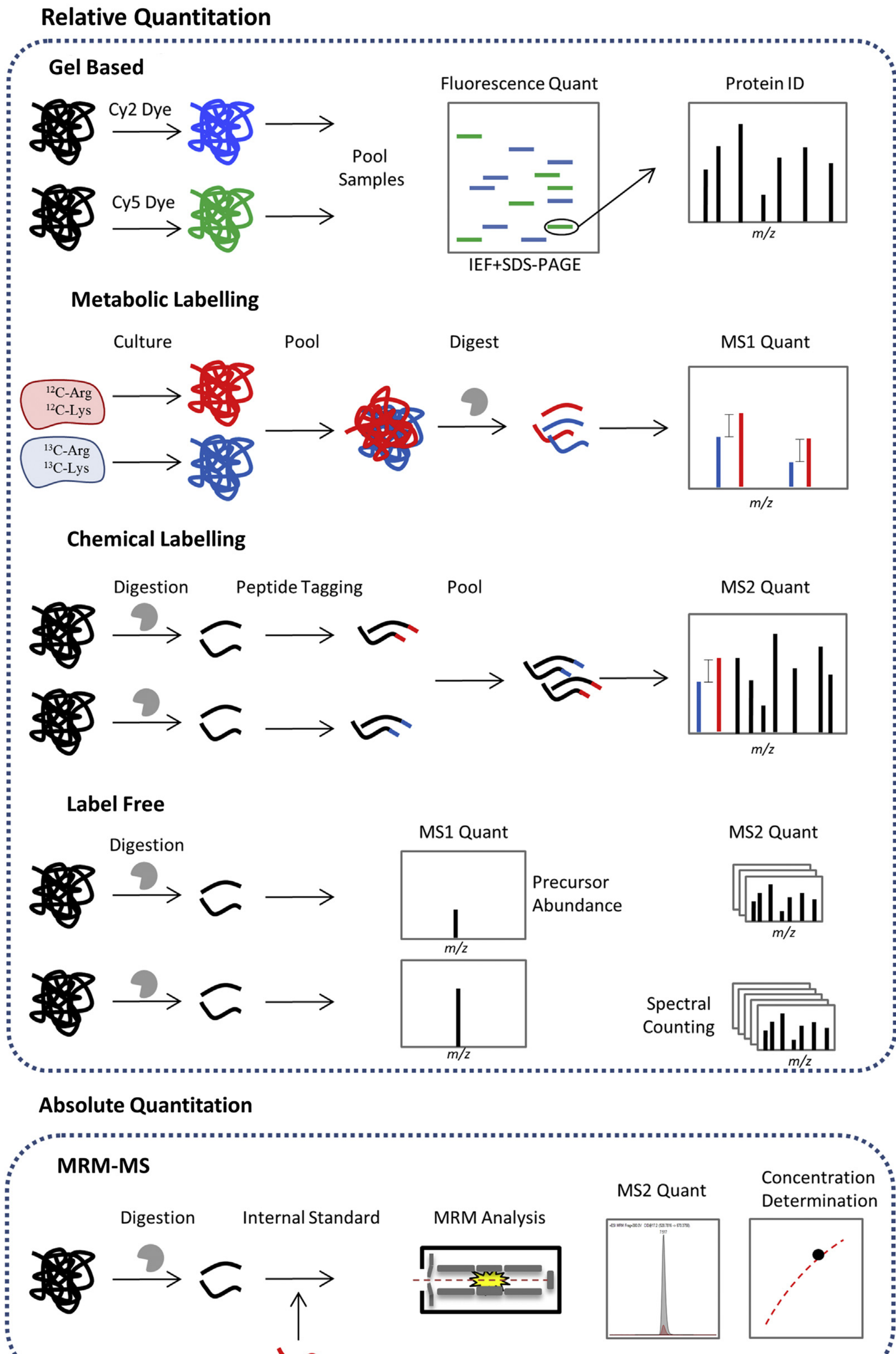


FIGURE 13.4 Overview of quantitative proteomics approaches. Relative quantitation (Quant) for biomarker discovery is classically performed using gel-based, metabolic, chemical, or label-free strategies. Absolute quantitation is achieved by integrating isotopically labeled internal standards with the multiple-reaction monitoring (MRM) technique on a triple quadrupole mass spectrometer. *MS*, mass spectrometry; *m/z*, mass-to-charge ratio.

interface to the mass spectrometer (Fournier et al., 2007; Wolters et al., 2001). Although these approaches have seen success, classical gel-based and offline/online fractionation approaches have largely been replaced by gel-free strategies.

Labeling Strategies: Two major quantitative strategies entail either metabolic or chemical labeling; the former uses stable isotopes whereas the latter generally uses isobaric tags. Stable isotope labeling with amino acids in cell culture (SILAC) is the predominant strategy used for metabolic labeling (Percy et al., 2012a). This is accomplished by growing cell lines in a medium containing heavy labeled isotopes of a specific amino acid (e.g., ^{15}N or ^{13}C lysine and arginine); the result is a set of peptides that, when cleaved using trypsin, contain at least one heavy labeled amino acid (Percy et al., 2012b). Labeled and nonlabeled peptides act identically during digestion, HPLC, and mass analysis; a measurable mass shift associated with their unique isotopic makeup allows quantitation (Ong and Mann, 2006). This is a popular approach but is limited to cells that can be cultured in the SILAC media; this renders many human and clinical samples (e.g., serum and tissues) incompatible. In addition, the strategy allows for only three different conditions: nonlabeled, labeled with ^{13}C-lysine/arginine, and labeled with ^{15}N- and ^{13}C-labeled lysine/arginine.

Chemical labeling using amine-reactive mass tags is the other predominant labeling methodology. Tandem mass tags (TMTs) and isobaric tags for relative and absolute quantitation (iTRAQ) both incorporate isobaric-stable isotope tags via a reactive N-hydroxy succinimide chemistry (Selevsek et al., 2015; Wu et al., 2016). Each tag incorporates a reporter group, balance or normalizing group, and reactive group that covalently bonds to tryptic peptides. As the name states, each individual tag will have the same mass whereas the individual reporter and balance group mass will vary. For example, a 4-plex assay contains four isobaric tags with reporter masses of 113, 114, 115, and 115 Da and balancing groups with masses of 32, 31, 30, and 29 Da. During mass spectral analysis, tags are released during peptide fragmentation and their relative abundance between samples is determined by integrating the areas of each individual reporter tags. Modern iTRAQ and TMT approaches have matured such that up to 10 samples can be quantified using the TMT approach, and up to 8 with iTRAQ (Schoenherr et al., 2015; Kuhn and Carr, 2016).

Label-Free Quantification: Label free−based quantitation has become a prevalent alternative to labeling strategies, with several advantages and some disadvantages. There are two fundamental approaches to label-free methodologies: quantification by signal intensity or spectral counting (Zhu et al., 2010). Using peptide signal intensity, extracted ion chromatograms (EICs) are generated for each distinct *m/z* ratio across the chromatographic separation. Each EIC is integrated and each peak area can be used for quantitative comparisons between disease and control samples. In contrast, spectral counting relies on fragmentation and database matching for quantitative comparisons. Put simply, the more abundant a protein is in the sample, the more fragmentation spectra will be generated, enabling relative quantification between samples. In the latter case, an accurate mass and retention time (AMRT) library is developed using MS/MS data. Samples are analyzed in MS-only mode and the peak area is compared across samples. Using differential analysis software, peaks that are shown to be statistically different are identified using the AMRT library. In general, nonlabeling strategies employ simple fractionation and sample preparation steps and also allow for the analysis of much larger sample sizes, as opposed to other approaches. Conversely, reliable quantitative measurements rely heavily on biological and sample replicates coupled with reproducible mass spectral analysis (Bantscheff et al., 2007).

13.3.4.2 Absolute Quantitation Strategies for Verification

After biomarker discovery, targeted approaches are employed to quantify specific biomarker candidates rapidly and reliably using a large cohort of clinical samples. This workflow universally employs a triple quadrupole (QQQ) mass spectrometer operating in MRM mode (Domon and Aebersold, 2010). These systems have been described unanimously as the reference standard of modern quantitative analysis for targeted small-molecule quantitation (Boyd et al., 2011); they are still being tested for their utility in quantitative proteomics. Using a QQQ system, quadrupole mass analyzers operate in tandem to isolate, fragment, and resolve specific peptide fragments. In MRM mode, two separate selection steps are performed. First, quadrupole 1 (Q1) isolates a select peptide (precursor ion), precursor ions are fragmented in collision cell q2, and then specific peptide fragment (product ions) are further isolated in quadrupole 3 (Q3). Thus, protein measurements are obtained by acquiring data on a highly specific peptide fragment. This twofold ion filtration approach greatly reduces background noise, resulting in the most selective and sensitive acquisition mode currently employed by protein scientists (Elschenbroich and Kislinger, 2011). Quantitation is achieved by incorporating custom-synthesized internal standards in conjunction with external calibration curves. Running in MRM mode, internal standards and the native digested peptide are monitored simultaneously in the instrument. Calculating the ratio between standard and native peptide allows for the absolute quantitation (concentration determination) of the peptide targets in the sample (Yocum and Chinnaiyan, 2009).

13.4 BIOINFORMATICS

The quantitative strategies and MS platforms described thus far are agnostic to the nature of the sample type or treatment group; methods for data generation are rapidly becoming standardized. Therefore, one major challenge lies in meaningful analysis and interpretation of data. Many variations can exist but a typical proteomics informatics pipeline includes identification, quantification, significance testing, and biological interpretation (Fig. 13.5). Identification of peptides and proteins by MS are often performed using the Trans-Proteomics Pipeline (TPP) (Kenar et al., 2014), a highly developed suite of tools. Within the pipeline are tools for peptide and protein identification (X! Tandem, Comet-MS, Mascot, Protein Prophet, and Protein Prophet), MS1 quantification (MSInspect), and functions for gene ontology analysis (Eng et al., 2013; Bellew et al., 2006; Keller et al., 2002; Nesvizhskii et al., 2003; Craig and Beavis, 2004).

Outside the TPP, a complement of tools is available that can be integrated with TPP results. These include tools for MS/MS quantitation by spectral counting (Abacus and Scaffold) (Searle, 2010; Fermin et al., 2011), significance analysis of MS1 or MS2 data (QProt) (Choi et al., 2015), pathways analysis (Cytoscape, Ingenuity) (Shannon et al., 2003), and a rich statistical environment (R, www.CRAN.org) for customized analysis and presentation of results.

Regardless of the pipeline tools used, proteomics workflows commonly adopt three strategies: (1) nontargeted methods to identify and semiquantify, (2) differential analysis by MS1 quantitation followed by targeted MS2 identification, and (3) a hybrid AMRT-based approach. The study of PTMs is extremely important in many studies; the TPP MS2 search engines support custom modification analysis for PTMs and also chemical modifications (e.g., biotinylated). The accuracy of these assigned PTMs can be measured (LuXor/LuciPhor) (Fermin et al., 2013) and can improve the confidence and improve quantitative accuracy when profiling peptides with PTMs.

As mentioned, biological interpretation is by far the most customized analysis requiring the most flexibility and collaboration. As first steps, ontologies analysis [e.g., DAVID (Huang Da et al., 2009) and/or Panther (Mi et al., 2016)], protein set enrichment [e.g., PSEA-Quant (Lavallee-Adam et al., 2014)], protein−protein interaction [BioGrid (Oughtred et al., 2016)], and pathways [Cytoscape, Ingenuity (Shannon et al., 2003)] analysis can be performed.

Validation of all peptide and protein quantitation can be performed by extracting ion chromatograms and performing *t* tests. Absolute quantitation methods using MRM are developed using empirical data and a Skyline workflow, with or without internal standards. Validation can also be performed using antibodies, when available.

13.5 EXAMPLES OF DISCOVERY AND VERIFICATION PROTEOMICS

To date, most proteomics-based studies have focused on the discovery of disease-specific protein biomarkers. These accomplishments have employed a variety of quantitative approaches from nearly all clinically relevant biological specimens. In addition, verification workflows have begun to identify a number of protein biomarkers successfully with clinical applicability. A small cohort of examples from both of these approaches can be found in Table 13.1.

13.6 EXAMPLES OF PROTEIN-BASED DIAGNOSTICS ASSAYS

The identification of protein biomarkers serves two critical roles in clinical medicine; they can be used to provide insight into the pathophysiology of disease and to guide clinical treatment options. While not necessarily using a proteomics-based approach, there are a multitude of examples of proteins being employed for medical diagnostics.

One such example is CRP, a plasma protein produced by the liver and indicative of an inflammatory process (Black et al., 2004). High levels of CRP (above 2 mg/L) in the blood have been correlated with an increased chance of a heart attack or stroke but can also indicate a bone

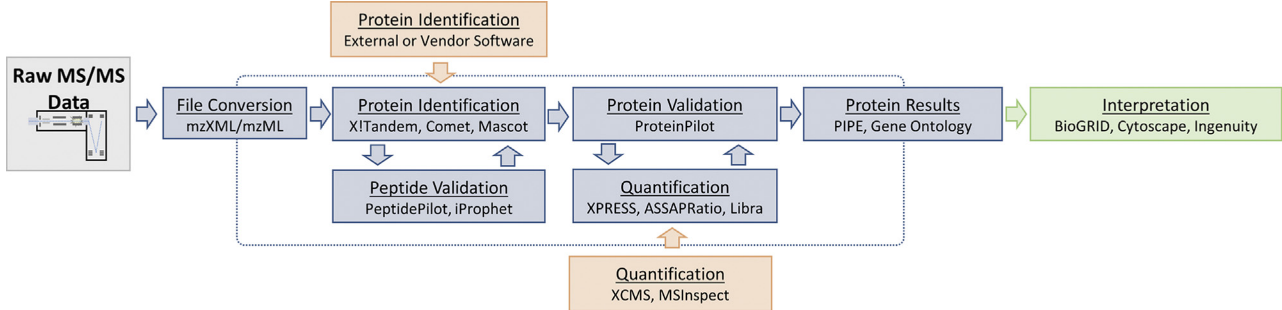


FIGURE 13.5 Proteomics informatics pipeline including tools for protein and peptide identification and validation, relative or absolute quantitation, statistical analysis, and biological and/or pathway interpretation. *MS/MS*, tandem mass spectrometry.

TABLE 13.1 Examples of Discovery and Verification Studies Including Disease Types, Sample Sources, and Analytical Approach

Approach	Disease	Study	Sample Type	Sample Preparation	Quantitation	References
Biomarker discovery	Alzheimer	Disease progression makers of Alzheimers disease	Cerebrospinal Fluid	Immunodepletion	Label-free	Hendrickson et al. (2015)
	Alzheimer	Identification of proteins that are differentially expressed in brains with Alzheimer's disease	Brain tissue	Isoelectric focusing + RP	iTRAQ	Minjarez et al. (2016)
	Ovarian cancer	Analysis of ovarian cancer tissue to identify biomarkers of responders and nonresponses	Human ovary tissue	–	2D-DIGE	Sehrawat et al. (2016)
	Lung cancer	Targets of oncogenic ECF receptor signaling in lung cancer	Cancer cells	SCX + RP; titanium dioxide	SILAC	Zhang et al. (2015)
	Breast cancer	Label-free LC-MS Analysis of HER2+ breast cancer cell line	Cancer cells	–	Label-free	Di Luca et al. (2015)
	Breast cancer	Biomarker discovery of laser capture microdissected breast cancer tissue	Tumor tissue	Laser microdissection	Label-free	Liu et al. (2012)
	Bladder cancer	Discovery of novel bladder cancer biomarkers in urine	Urine	SCX + RP	iTRAQ	Chen et al. (2010)
	Crohn disease	Serum protein profiling of adults and children with Crohn disease	Serum	–	2D-DIGE	Vaiopoulou et al. (2015)
	COPD	Proteomic analysis in lung tissue of smokers and COPD patients	Lung tissue	–	2D-DIGE	Lee et al. (2009)
	COPD	Sputum proteomics identifies elevated levels in smokers and COPD	Sputum and plasma	–	2D-DIGE	Ohlmeier et al. (2012)
Biomarker verification	Lung cancer	Verification of the biomarker candidates for non–small-cell lung cancer	Plasma	Immunodepletion	Absolute	Kim et al. (2015)
	Lung cancer	Verification of serological biomarker candidates for lung adenocarcinoma	Whole blood	Immunodepletion	Absolute	Wu et al. (2015)
	Breast cancer	Validation of a novel plasma protein signature for breast cancer diagnosis	Plasma	–	Absolute	Lee et al. (2015)
	Breast cancer	Proteomic analyses of decorin and endoplasmin (HSP90B1) are associated with breast cancer metastasis	Cancer tissue	–	iTRAQ and absolute	Cawthorn et al. (2012)
	Pancreatic cancer	Novel protein markers of pancreatic cancer using formalin-fixed paraffin-embedded tissues	Pancreatic tissue	–	Absolute and IHC	Takadate et al. (2013)
	Pancreatic cancer	Targeted proteomic assay for pancreatic cancer biomarkers	Plasma	Immunodepletion	Absolute	Pan et al. (2012)
	Alzheimer	MS and flow sorting analysis to detect the decreased serum apolipoprotein	Serum	Cell sorting	Absolute	Han et al. (2014)

2D-DIGE, two-dimensional difference gel electrophoresis; *COPD*, chronic obstructive pulmonary disease; *IHC*, Immunohistochemistry; *iTRAQ*, isobaric tag for relative and absolute quantitation; *RP*, reversed-phase; *SCX*, strong cation exchange; *SILAC*, stable isotope labeling with amino acids in cell culture.

infection, arthritic flare-up, inflammatory bowel disease, tuberculosis, autoimmune disease, cancer, or pneumonia (Danesh et al., 2004). Higher CRP levels have also been detected in women during the second half of pregnancy and in women receiving hormonal contraceptives (Shetlar et al., 1955). Therefore, this test is most useful when it is used in conjunction with further diagnostic testing.

A more specific test for the diagnostic workup and management of acute coronary syndromes (ACS) is the measurement of cardiac troponin (cTn) levels. Troponins include three proteins (troponins C, I, and T) that regulate actin and myosin during muscle contractions. The latter two, cardiac troponin I (cTnI) and cardiac troponin T (cTnT), have distinct isoforms that are unique to cardiac muscle and are released into the bloodstream during cardiomyocyte necrosis (Mahajan and Jarolim, 2011; Wang, 2007). These protein isoforms serve as the preferred diagnostic test for ACS because of their tissue-specific expression. Advances in immunoassay technologies have allowed for increased sensitivity for cTn, lowering the potential for a missed diagnosis of ACS (Saenger et al., 2011). However, this comes at the cost of specificity, allowing for the detection of low levels of cTn even in samples collected from healthy subjects. This enhancement in sensitivity has made it critical to define a concentration limit at which cTn concentrations will be reported as positive and indicative of myocardial necrosis suggestive of ACS.

B-type natriuretic peptide (BNP) and N-terminal-pro-BNP (NT-pro-BNP) are polypeptides secreted by the cardiac ventricles in response to ventricular wall tension as a result of volume expansion (Maisel et al., 2002). The harder the heart has to work, such as during heart failure, the more BNP is released (McCullough et al., 2002). Blood BNP levels can therefore be used to detect and diagnose heart failure as well as to demonstrate the patient's response to heart failure treatment (Dao et al., 2001; Davis et al., 1994). BNP and NT-pro-BNP levels may also be used to evaluate a patient's risk for future cardiac events, because increased levels in individuals with ACS have been shown to correlate with an increased risk of recurrent events. As with cTn testing, a cutoff value must be established to predict congestive heart failure accurately (Maisel et al., 2002).

Another example of protein detection used for clinical diagnostics is the detection of PSA for the screening, assessment of treatment, and diagnosis of prostate cancer (Stamey et al., 1987). Before PSA testing, there had to be a palpable abnormality of the prostate before a biopsy would be performed, at which point most patients had metastatic disease. Since PSA evaluation, the chances of metastases at the time of diagnosis have significantly decreased. However, PSA levels may also rise in men with benign prostatic hyperplasia. Therefore, studies are under way to determine definitively whether PSA screening makes a significant difference in the detection of prostate cancer, and its subsequent survival rate.

The lack of specificity of these tests makes them good screening tools when they are used in conjunction with other diagnostic measures. The added sensitivity of testing techniques has only worked to decrease the specificity of these assays. However, when a panel of biomarkers can be assessed instead of a single protein, the specificity of the assay may be increased. Such is the case with OVA1 testing by Vermillion, Inc., which uses five biomarkers to diagnose ovarian cancer. Historically, the serum biomarker cancer antigen 125 (CA125) was used as a diagnostic tool for malignancy in women with pelvic masses (Visintin et al., 2008). Elevated CA125 levels could be detected in women with metastatic epithelial ovarian cancer; however, a number of benign conditions could also result in elevated levels (Bast et al., 1997). A proteomics biomarker discovery approach led to the assessment of additional markers: transthyretin, apolipoprotein, A-1, 2-microglobulin, and transferrin, in the prediction of ovarian cancer. These biomarkers, together with CA125, have been incorporated into the multivariate immunoassay known as OVA1, the use of which has been approved by the US FDA in women with pelvic masses (FDA, 2009). This is the only currently FDA-approved diagnostic test based on proteomics biomarker discovery. According to studies, OVA1 is more sensitive in detecting ovarian cancers than either clinician assessment or CA125 alone. Compared with clinician assessment, the use of OVA1 in a diagnostic workup was able to identify 70% of cancers correctly that had been missed by gynecologists and 95% of cancers that had been missed by gynecologic oncologists. Compared with the use of CA125 alone in diagnosing stage I disease, the sensitivity for OVA1 was 90% whereas that of CA125 was 61% (Ueland et al., 2011).

Another proteomics panel used as a blood or serum-based diagnostic tool for patients with non−small cell lung cancer (NSCLC) is VeriStrat. This is a commercially available test based on MALDI-MS analysis. This assay was developed after testing serum samples from patients with a diagnosis of NSCLC who experienced long-term stable disease or early progression with gefitinib therapy (Frasci et al., 2000). Eight mass-spectral features were identified as differentiating between these groups when measuring for several acute-phase reactant proteins in the blood. An algorithm was created to assign each spectrum a binary classification of good or poor correlating with more or less favorable prognosis (Stinchcombe et al., 2013).

Through shotgun proteomics, thousands of proteins in complex samples can be assessed for use in proteomics-based diagnostic testing. Such an approach was used by the Mayo Clinic, which developed a diagnostic tool that can subtype amyloidosis proteins, a disease characterized by abnormal deposits of cellular proteins in extracellular

sites resulting in organ damage. Shotgun proteomics was used to identify amyloid proteins from microdissected biopsy tissue, allowing for the characterization of more than 20 subtypes of amyloidosis. The underlying cause and treatment of each subtype of amyloidosis can be radically different, requiring accurate subtype identification. This diagnostic method combines proteomics MS analysis with laser microdissection after cell staining to identify amyloid type with 100% sensitivity and 40−80% specificity (Theis et al., 2013).

Microbiologic diagnostics aims to identify the causative agents of infectious disease such as bacteria, fungi, and parasites. Historically, diagnosis relied on microscopic techniques of cultured specimens; however, rapid advances in the species identification of microorganisms through the use of MALDI-TOF MS reduced the time and cost of diagnosis with demonstrated reliability and accuracy. Organic solvents are used to extract proteins from cultured microorganisms. These proteins, mainly ribosomal proteins present in high concentrations, are then analyzed using MALDI-TOF MS. Pathogen identification is based on the analysis of ribosomal protein sequences, allowing for species identification from a colony and communication with the responsible physician within a few minutes (Wieser et al., 2012). Launched in 2006, BioMérieux's Vitek MS clinical microbiology MS system was the first platform of its kind cleared for clinical use in the United States by the US FDA; the Vitek MS platform was cleared in 2011. Both MALDI-TOF MS-based platforms identify microorganisms by matching ribosomal protein profiles of samples with those contained in a proprietary database (Jarman et al., 2000). Consequently, species that do not contain sufficient differences in their ribosomal protein sequence cannot be distinguished with this technique. Current research aims to detect bacterial resistance mechanisms using MALDI-TOF MS (Wieser et al., 2012).

13.7 CHALLENGES IN CLINICAL PROTEOMICS

Although it is clearly essential and important for the advancement of clinical biomarkers, the field of clinical proteomics is faced with challenges. As mentioned, this includes highly abundant plasma proteins, limits in the sample numbers that can be labeled, and technical reproducibility across large batches. Additional challenges include sample availability, sample collection and handling, sample numbers, and cost. For example, although formalin-fixed and paraffin-embedded tissue is widely available, issues with extraction and depth of coverage make it a poor choice for most discovery approaches. Because of the high variability in human samples, a very large sample size is usually required to provide statistical power; however, with sample preparation, data acquisition, and analysis requiring upward of 20 h and thousands of dollars per sample, sample size is almost always limited. Specific to quantitative proteomics, reproducibility and sensitivity are additional issues. Finally, sample and analysis throughput and data interpretation are also challenges. Despite the availability of sophisticated pathway analysis programs, a thorough interpretation has often been beyond the scope of many research programs.

13.8 FUTURE ADVANCES AND CONCLUDING REMARKS

Clearly, there is immense potential for generating protein biomarkers using clinical proteomics studies. Quantitative proteomics techniques have matured but many challenges remain. A multitude of advances are being implemented to increase sample throughput, decrease per-sample cost, and accelerate method development timelines, while increasing assay sensitivity to detect even the lowest abundant biomarker targets.

Arguably, a principal concern with transitioning biomarker candidates is laborious sample-handling requirements and lengthy time requirements on the MS platforms. To expedite sample handling, several automation solutions have been designed to batch process multiple microwell plates in parallel (Selevsek et al., 2015; Wu et al., 2016). In conjunction with sample preparation, migration toward high-flow separation systems, or ultraperformance liquid chromatography (UPLC) has also seen success. Using a UPLC, run times can be reduced to less than 10 min/sample as opposed to more than 60 min/sample with conventional nanoflow systems, which is critical for the throughput demands in biomarker verification studies.

Advances in instrumentation, acquisition methods, and software workflow also show promise in generating or monitoring specific protein targets. One novel acquisition approach using a quadrupole TOF-MS is sequential window acquisition of all theoretical fragment ion spectra (SWATH). This acquisition approach combines high-resolution MS with pseudo-MRM quantitation. Using SWATH, precursor ion windows (e.g., 400−425 *m/z*) are isolated in a quadrupole mass filter and fragmented, and a high-resolution spectrum is generated for all peptides within the mass isolation window. Quantitation is achieved by integrating fragment ion abundances. This unique acquisition method has the unique advantage of providing near-global coverage of all peptides in a sample versus highly specific MRM workflow (Liu et al., 2013). Another approach, parallel reaction monitoring, employs an upfront quadrupole mass filter combined with a high-resolution orbitrap mass analyzer, enabling accurate, high-resolution,

and interference-free product ion measurements (Majovsky et al., 2014). In conjunction with novel hardware, a multitude of tools continue to coevolve to handle data from novel acquisition methods while providing new tools for traditional discovery and validation workflow (Röst et al., 2014). For example, it is now reasonably straightforward to develop an entire targeted assay in silico using online libraries and databases such as the SRMAtlas/PeptideAtlas (Kusebauch et al., 2014).

In terms of assay sensitivity, removal of highly abundant plasma proteins via immunoaffinity chromatography is a routine technique for detecting and quantifying low-abundant protein biomarkers. This approach is a routine technique during both biomarker discovery and verification. More recently, postdigestion antipeptide immunoenrichment workflows have become increasingly prevalent. The workflow, originally is called stable isotope standard capture with antipeptide antibodies (SISCAPA), has also been named immuno−multiple reaction monitoring (iMRM) (Anderson et al., 2004a; De Marchi et al., 2015). In these approaches, synthetic peptides are conjugated to a carrier protein and used to generate peptide-specific antibodies from a host animal, which in turn is used to purify selected peptides from a proteolytic background (Wu et al., 2016). These approaches have been highly successful in increasing detection limits by more than an order of magnitude. The iMRM/SISCAPA workflow is also amenable to automation and ultrafast analysis; one study required only 8 seconds per sample on an Agilent RapidFire system (Razavi et al., 2012).

In addition, by further integrating protein liquid chromatography/MS/MS discovery and verification workflows in the clinical space, there is the potential to develop peptide and protein-based clinical assays as opposed to traditional antibody-based approaches. Although unlikely to supplant traditional ligand-binding techniques, MS-based approaches offer several advantages over traditional antibody-based assays. Advantages include greater specificity as MS-based approaches are not subject to antibody cross-reactivity, improvements in cost- and time- of method development, the ability to monitor protein isoforms or PTMs reliably, and finally the ability to multiplex multiple peptide targets into a single assay.

This chapter has described the state of the art of current proteomics-based approaches to clinical medicine. This includes biomarker discovery and validation, several examples of applied proteomics, and finally, several next-generation technologies that it is hoped will accelerate clinical implementation. Although the road from discovery to clinical application has been slow, there have been tremendous efforts made in past few decades in translating protein biomarkers from candidate to bona fide disease indicators. There is clearly reason to be optimistic with several assays already online and groundwork laid for FDA approval of additional tests. These continued efforts will clearly be beneficial for both clinicians and patients in the near future.

REFERENCES

Alban, A., David, S.O., Bjorkesten, L., Andersson, C., Sloge, E., Lewis, S., Currie, I., 2003. A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. Proteomics 3, 36−44.

Anderson, L., 2005. Candidate-based proteomics in the search for biomarkers of cardiovascular disease. J. Physiol. 563, 23−60.

Anderson, N.L., Anderson, N.G., Haines, L.R., Hardie, D.B., Olafson, R.W., Pearson, T.W., 2004a. Mass spectrometric quantitation of peptides and proteins using stable isotope standards and capture by anti-peptide antibodies (SISCAPA). J. Proteome Res. 3, 235−244.

Anderson, N.L., Polanski, M., Pieper, R., Gatlin, T., Tirumalai, R.S., Conrads, T.P., Veenstra, T.D., Adkins, J.N., Pounds, J.G., Fagan, R., 2004b. The human plasma proteome. Mol. Cell Proteomics 3, 311−326.

Baker, M.W., Atkins, A.E., Cordovado, S.K., Hendrix, M., Earley, M.C., Farrell, P.M., 2016. Improving newborn screening for cystic fibrosis using next-generation sequencing technology: a technical feasibility study. Genet. Med. 18, 231−238.

Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., Kuster, B., 2007. Quantitative mass spectrometry in proteomics: a critical review. Anal. Bioanal. Chem. 389, 1017−1031.

Bast Jr., R., Xu, F., Yu, Y., Barnhill, S., Zhang, Z., Mills, G., 1997. CA 125: the past and the future. Int. J. Biol. Markers 13, 179−187.

Beck, M., Schmidt, A., Malmstroem, J., Claassen, M., Ori, A., Szymborska, A., Herzog, F., Rinner, O., Ellenberg, J., Aebersold, R., 2011. The quantitative proteome of a human cell line. Mol. Syst. Biol. 7, 549.

Bellew, M., Coram, M., Fitzgibbon, M., Igra, M., Randolph, T., Wang, P., May, D., Eng, J., Fang, R., Lin, C., Chen, J., Goodlett, D., Whiteaker, J., Paulovich, A., Mcintosh, M., 2006. A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution LC-MS. Bioinformatics 22, 1902−1909.

Bennett, M.J., 2014. Newborn screening for metabolic diseases: saving children's lives and improving outcomes. Clin. Biochem. 47, 693−694.

Black, S., Kushner, I., Samols, D., 2004. C-reactive protein. J. Biol. Chem. 279, 48487−48490.

Boyd, R.K., Basic, C., Bethem, R.A., 2011. Trace Quantitative Analysis by Mass Spectrometry. Wiley.

Cancer, C.P.T.F., 2009. Advancing Protein Science for Personalized Medicine 2008−09 Annual Report National Institutes of Health.

Cappellini, E., Jensen, L.J., Szklarczyk, D., Ginolhac, A., Da Fonseca, R.A., Stafford Jr., T.W., Holen, S.R., Collins, M.J., Orlando, L., Willerslev, E., 2011. Proteomic analysis of a pleistocene mammoth femur reveals more than one hundred ancient bone proteins. J. Proteome Res. 11, 917−926.

Cargile, B.J., Sevinsky, J.R., Essader, A.S., Stephenson, J.L., Bundy, J.L., 2005. Immobilized pH gradient isoelectric focusing as a first-dimension separation in shotgun proteomics. J. Biomol. Tech. JBT 16, 181−189.

Cawthorn, T.R., Moreno, J.C., Dharsee, M., Tran-Thanh, D., Ackloo, S., Zhu, P.H., Sardana, G., Chen, J., Kupchak, P., Jacks, L.M., Miller, N.A., Youngson, B.J., Iakovlev, V., Guidos, C.J., Vallis, K.A., Evans, K.R., Mccready, D., Leong, W.L., Done, S.J., 2012. Proteomic analyses reveal high expression of decorin and endoplasmin (HSP90B1) are associated with breast cancer metastasis and decreased survival. PLoS One 7, e30992.

Chaerkady, R., Harsha, H.C., Nalli, A., Gucek, M., Vivekanandan, P., Akhtar, J., Cole, R.N., Simmers, J., Schulick, R.D., Singh, S., Torbenson, M., Pandey, A., Thuluvath, P.J., 2008. A quantitative proteomic approach for identification of potential biomarkers in hepatocellular carcinoma. J. Proteome Res. 7, 4289−4298.

Chen, Y.T., Chen, C.L., Chen, H.W., Chung, T., Wu, C.C., Chen, C.D., Hsu, C.W., Chen, M.C., Tsui, K.H., Chang, P.L., Chang, Y.S., Yu, J.S., 2010. Discovery of novel bladder cancer biomarkers by comparative urine proteomics using iTRAQ technology. J. Proteome Res. 9, 5803−5815.

Choi, H., Kim, S., Fermin, D., Tsou, C.C., Nesvizhskii, A.I., 2015. QPROT: statistical method for testing differential expression using protein-level intensity data in label-free quantitative proteomics. J. Proteomics 129, 121−126.

Convention, U.S.P., 2015. Chromatography.

Craig, R., Beavis, R.C., 2004. TANDEM: matching proteins with tandem mass spectra. Bioinformatics 20, 1466−1467.

Danesh, J., Wheeler, J.G., Hirschfield, G.M., Eda, S., Eiriksdottir, G., Rumley, A., Lowe, G.D., Pepys, M.B., Gudnason, V., 2004. C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. N. Engl. J. Med. 350, 1387−1397.

Dao, Q., Krishnaswamy, P., Kazanegra, R., Harrison, A., Amirnovin, R., Lenert, L., Clopton, P., Alberto, J., Hlavin, P., Maisel, A.S., 2001. Utility of B-type natriuretic peptide in the diagnosis of congestive heart failure in an urgent-care setting. J. Am. Coll. Cardiol. 37, 379−385.

Davis, M., Espiner, E., Yandle, T., Richards, G., Town, I., Neill, A., Drennan, C., Richards, A., Turner, J., Billings, J., 1994. Plasma brain natriuretic peptide in assessment of acute dyspnoea. Lancet 343, 440−444.

De Marchi, T., Kuhn, E., Carr, S.A., Umar, A., 2015. Antibody-based capture of target peptides in multiple reaction monitoring experiments. Methods Mol. Biol. 1293, 123−135.

Den, R.B., Yousefi, K., Trabulsi, E.J., Abdollah, F., Choeurng, V., Feng, F.Y., Dicker, A.P., Lallas, C.D., Gomella, L.G., Davicioni, E., Karnes, R.J., 2015. Genomic classifier identifies men with adverse pathology after radical prostatectomy who benefit from adjuvant radiation therapy. J. Clin. Oncol. 33, 944−951.

Di Luca, A., Henry, M., Meleady, P., O'connor, R., 2015. Label-free LC-MS analysis of HER2+ breast cancer cell line response to HER2 inhibitor treatment. Daru 23, 40.

Domon, B., Aebersold, R., 2006. Mass spectrometry and protein analysis. Science 312, 212−217.

Domon, B., Aebersold, R., 2010. Options and considerations when selecting a quantitative proteomics strategy. Nat. Biotechnol. 28, 710−721.

Elschenbroich, S., Kislinger, T., 2011. Targeted proteomics by selected reaction monitoring mass spectrometry: applications to systems biology and biomarker discovery. Mol. Biosyst. 7, 292−303.

Eng, J.K., Jahan, T.A., Hoopmann, M.R., 2013. Comet: an open-source MS/MS sequence database search tool. Proteomics 13, 22−24.

FDA, U, 2009. FDA clears a test for ovarian cancer: test can help identify potential malignancies, guide surgical decisions. FDA News Release 11.

Fermin, D., Basrur, V., Yocum, A.K., Nesvizhskii, A.I., 2011. Abacus: a computational tool for extracting and pre-processing spectral count data for label-free quantitative proteomic analysis. Proteomics 11, 1340−1345.

Fermin, D., Walmsley, S.J., Gingras, A.C., Choi, H., Nesvizhskii, A.I., 2013. LuciPHOr: algorithm for phosphorylation site localization with false localization rate estimation using modified target-decoy approach. Mol. Cell Proteomics 12, 3409−3419.

Fournier, M.L., Gilmore, J.M., Martin-Brown, S.A., Washburn, M.P., 2007. Multidimensional separations-based shotgun proteomics. Chem. Rev. 107, 3654−3686.

Frantzi, M., Bhat, A., Latosinska, A., 2014. Clinical proteomic biomarkers: relevant issues on study design & technical considerations in biomarker development. Clin. Transl. Med. 3, 7.

Frasci, G., Lorusso, V., Panza, N., Comella, P., Nicolella, G., Bianco, A., De Cataldis, G., Iannelli, A., Bilancia, D., Belli, M., 2000. Gemcitabine plus vinorelbine versus vinorelbine alone in elderly patients with advanced non−small-cell lung cancer. J. Clin. Oncol. 18, 2529−2536.

Fröhlich, T., Arnold, G.J., 2009. A newcomer's guide to nano-liquid-chromatography of peptides. Proteomics Methods Protoc. 123−141.

Glish, G.L., Vachet, R.W., 2003. The basics of mass spectrometry in the twenty-first century. Nat. Rev. Drug Discov. 2, 140−150.

Gray, R.D., Macgregor, G., Noble, D., Imrie, M., Dewar, M., Boyd, A.C., Innes, J.A., Porteous, D.J., Greening, A.P., 2008. Sputum proteomics in inflammatory and suppurative respiratory diseases. Am. J. Respir. Crit. Care Med. 178, 444−452.

Hakimi, A., Auluck, J., Jones, G.D., Ng, L.L., Jones, D.J., 2014. Assessment of reproducibility in depletion and enrichment workflows for plasma proteomics using label-free quantitative data-independent LC-MS. Proteomics 14, 4−13.

Han, S.H., Kim, J.S., Lee, Y., Choi, H., Kim, J.W., Na, D.L., Yang, E.G., Yu, M.H., Hwang, D., Lee, C., Mook-Jung, I., 2014. Both targeted mass spectrometry and flow sorting analysis methods detected the decreased serum apolipoprotein E level in Alzheimer's disease patients. Mol. Cell Proteomics 13, 407−419.

Hardt, M., Thomas, L.R., Dixon, S.E., Newport, G., Agabian, N., Prakobphol, A., Hall, S.C., Witkowska, H.E., Fisher, S.J., 2005. Toward defining the human parotid gland salivary proteome and peptidome: identification and characterization using 2D SDS-PAGE, ultrafiltration, HPLC, and mass spectrometry. Biochemistry 44, 2885−2899.

Hendrickson, R.C., Lee, A.Y., Song, Q., Liaw, A., Wiener, M., Paweletz, C.P., Seeburger, J.L., Li, J., Meng, F., Deyanova, E.G., 2015. High resolution discovery proteomics reveals candidate disease progression markers of Alzheimer's disease in human cerebrospinal fluid. PloS One 10, e0135365.

Herbert, C.G., Johnstone, R.A., 2002. Mass Spectrometry Basics. CRC Press.

Hillenkamp, F., Karas, M., Holtkamp, D., Klüsener, P., 1986. Energy deposition in ultraviolet laser desorption mass spectrometry of biomolecules. Int. J. Mass Spectrom. Ion Process. 69, 265−276.

Honeyman, J.N., Simon, E.P., Robine, N., Chiaroni-Clarke, R., Darcy, D.G., Lim, I.I.P., Gleason, C.E., Murphy, J.M., Rosenberg, B.R., Teegan, L., 2014. Detection of a recurrent DNAJB1-

PRKACA chimeric transcript in fibrolamellar hepatocellular carcinoma. Science 343, 1010−1014.

Huang Da, W., Sherman, B.T., Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44−57.

Jarman, K.H., Cebula, S.T., Saenz, A.J., Petersen, C.E., Valentine, N.B., Kingsley, M.T., Wahl, K.L., 2000. An algorithm for automated bacterial identification using matrix-assisted laser desorption/ionization mass spectrometry. Anal. Chem. 72, 1217−1223.

Kebarle, P., Verkerk, U.H., 2009. Electrospray: from ions in solution to ions in the gas phase, what we know now. Mass Spectrom. Rev. 28, 898−917.

Keller, A., Nesvizhskii, A.I., Kolker, E., Aebersold, R., 2002. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal. Chem. 74, 5383−5392.

Kenar, E., Franken, H., Forcisi, S., Wormann, K., Haring, H.U., Lehmann, R., Schmitt-Kopplin, P., Zell, A., Kohlbacher, O., 2014. Automated label-free quantification of metabolites from liquid chromatography-mass spectrometry data. Mol. Cell Proteomics 13, 348−359.

Kim, Y.J., Sertamo, K., Pierrard, M.A., Mesmin, C., Kim, S.Y., Schlesser, M., Berchem, G., Domon, B., 2015. Verification of the biomarker candidates for non-small-cell lung cancer using a targeted proteomics approach. J. Proteome Res. 14, 1412−1419.

Kuhn, E., Carr, S.A., 2016. Multiplexed immunoaffinity enrichment of peptides with anti-peptide antibodies and quantification by stable isotope dilution multiple reaction monitoring mass spectrometry. Quantitative Proteomics by Mass Spectrometry, pp. 135−167.

Kusebauch, U., Deutsch, E.W., Campbell, D.S., Sun, Z., Farrah, T., Moritz, R.L., 2014. Using PeptideAtlas, SRMAtlas, and PASSEL: comprehensive resources for discovery and targeted proteomics. Curr. Protoc. Bioinf. 13.25. 1−13.25. 28.

Lavallee-Adam, M., Rauniyar, N., Mcclatchy, D.B., Yates 3RD, J.R., 2014. PSEA-Quant: a protein set enrichment analysis on label-free and label-based protein quantification data. J. Proteome Res. 13, 5496−5509.

Lee, E.J., In, K.H., Kim, J.H., Lee, S.Y., Shin, C., Shim, J.J., Kang, K.H., Yoo, S.H., Kim, C.H., Kim, H.K., Lee, S.H., Uhm, C.S., 2009. Proteomic analysis in lung tissue of smokers and COPD patients. Chest 135, 344−352.

Lee, H.B., Kang, U.B., Moon, H.G., Lee, J., Lee, K.M., Yi, M., Park, Y.S., Lee, J.W., Yu, J.H., Choi, S.H., Cho, S.H., Lee, C., Han, W., Noh, D.Y., 2015. Development and validation of a novel plasma protein signature for breast cancer diagnosis by using multiple reaction monitoring-based mass spectrometry. Anticancer Res. 35, 6271−6279.

Lewis, J.K., Wei, J., Siuzdak, G., 2000. Matrix-assisted laser desorption/ionization mass spectrometry in peptide and protein analysis. Encycl. Anal. Chem. 5880−5894.

Li, D., Chan, D.W., 2014. Proteomic cancer biomarkers from discovery to approval: it's worth the effort. Expert Rev. Proteomics 11, 135−136.

Lindgren, J., Uvdal, P., Engdahl, A., Lee, A.H., Alwmark, C., Bergquist, K.-E., Nilsson, E., Ekström, P., Rasmussen, M., Douglas, D.A., 2011. Microspectroscopic evidence of Cretaceous bone proteins. PLoS One 6, e19445.

Liu, H., Lin, D., Yates 3RD, J.R., 2002. Multidimensional separations for protein/peptide analysis in the post-genomic era. Biotechniques 32, 898, 900, 902 passim.

Liu, N.Q., Braakman, R.B., Stingl, C., Luider, T.M., Martens, J.W., Foekens, J.A., Umar, A., 2012. Proteomics pipeline for biomarker discovery of laser capture microdissected breast cancer tissue. J. Mammary Gland. Biol. Neoplasia 17, 155−164.

Liu, Y., Huettenhain, R., Surinova, S., Gillet, L.C., Mouritsen, J., Brunner, R., Navarro, P., Aebersold, R., 2013. Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS. Proteomics 13, 1247−1256.

Magi, B., Bargagli, E., Bini, L., Rottoli, P., 2006. Proteome analysis of bronchoalveolar lavage in lung diseases. Proteomics 6, 6354−6369.

Mahajan, V.S., Jarolim, P., 2011. How to interpret elevated cardiac troponin levels. Circulation 124, 2350−2354.

Maisel, A.S., Krishnaswamy, P., Nowak, R.M., Mccord, J., Hollander, J.E., Duc, P., Omland, T., Storrow, A.B., Abraham, W.T., Wu, A.H., 2002. Rapid measurement of B-type natriuretic peptide in the emergency diagnosis of heart failure. N. Engl. J. Med. 347, 161−167.

Majovsky, P., Naumann, C., Lee, C.-W., Lassowskat, I., Trujillo, M., Dissmeyer, N., Hoehenwarter, W., 2014. Targeted proteomics analysis of protein degradation in plant signaling on an LTQ-Orbitrap mass spectrometer. J. Proteome Res. 13, 4246−4258.

Mccullough, P.A., Nowak, R.M., Mccord, J., Hollander, J.E., Herrmann, H.C., Steg, P.G., Duc, P., Westheim, A., Omland, T., Knudsen, C.W., 2002. B-type natriuretic peptide and clinical judgment in emergency diagnosis of heart failure analysis from breathing not properly (BNP) multinational study. Circulation 106, 416−422.

Mi, H., Poudel, S., Muruganujan, A., Casagrande, J.T., Thomas, P.D., 2016. PANTHER version 10: expanded protein families and functions, and analysis tools. Nucleic Acids Res. 44, D336−D342.

Minjarez, B., Calderon-Gonzalez, K.G., Rustarazo, M.L., Herrera-Aguirre, M.E., Labra-Barrios, M.L., Rincon-Limas, D.E., Del Pino, M.M., Mena, R., Luna-Arias, J.P., 2016. Identification of proteins that are differentially expressed in brains with Alzheimer's disease using iTRAQ labeling and tandem mass spectrometry. J. Proteomics 139, 103−121.

Nagaraj, N., Wisniewski, J.R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Paabo, S., Mann, M., 2011. Deep proteome and transcriptome mapping of a human cancer cell line. Mol. Syst. Biol. 7, 548.

Nesvizhskii, A.I., Keller, A., Kolker, E., Aebersold, R., 2003. A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. 75, 4646−4658.

Nicholas, B., Skipp, P., Mould, R., Rennard, S., Davies, D.E., O'connor, C.D., Djukanovic, R., 2006. Shotgun proteomic analysis of human-induced sputum. Proteomics 6, 4390−4401.

Ohlmeier, S., Mazur, W., Linja-Aho, A., Louhelainen, N., Ronty, M., Toljamo, T., Bergmann, U., Kinnula, V.L., 2012. Sputum proteomics identifies elevated PIGR levels in smokers and mild-to-moderate COPD. J. Proteome Res. 11, 599−608.

Ong, S.E., Mann, M., 2006. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). Nat. Protoc. 1, 2650−2660.

Oughtred, R., Chatr-Aryamontri, A., Breitkreutz, B.J., Chang, C.S., Rust, J.M., Theesfeld, C.L., Heinicke, S., Breitkreutz, A., Chen, D., Hirschman, J., Kolas, N., Livstone, M.S., Nixon, J., O'donnell, L., Ramage, L., Winter, A., Reguly, T., Sellam, A., Stark, C., Boucher, L., Dolinski, K., Tyers, M., January 4, 2016. BioGRID: a resource for studying biological interactions in yeast. Cold Spring Harb. Protoc. 2016 (1).

Pan, S., Chen, R., Brand, R.E., Hawley, S., Tamura, Y., Gafken, P.R., Milless, B.P., Goodlett, D.R., Rush, J., Brentnall, T.A., 2012. Multiplex targeted proteomic assay for biomarker detection in plasma: a pancreatic cancer biomarker case study. J. Proteome Res. 11, 1937−1948.

Patel, B.B., Barrero, C.A., Braverman, A., Kim, P.D., Jones, K.A., Chen, D.E., Bowler, R.P., Merali, S., Kelsen, S.G., Yeung, A.T., 2012. Assessment of two immunodepletion methods: off-target effects and variations in immunodepletion efficiency may confound plasma proteomics. J. Proteome Res. 11, 5947−5958.

Peng, J., Elias, J.E., Thoreen, C.C., Licklider, L.J., Gygi, S.P., 2003. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J. Proteome Res. 2, 43−50.

Percy, A.J., Chambers, A.G., Smith, D.S., Borchers, C.H., 2012a. Standardized protocols for quality control of MRM-based plasma proteomic workflows. J. Proteome Res. 12, 222−233.

Percy, A.J., Chambers, A.G., Yang, J., Borchers, C.H., 2013. Multiplexed MRM-based quantitation of candidate cancer biomarker proteins in undepleted and non-enriched human plasma. Proteomics 13, 2202−2215.

Percy, A.J., Chambers, A.G., Yang, J., Domanski, D., Borchers, C.H., 2012b. Comparison of standard-and nano-flow liquid chromatography platforms for MRM-based quantitation of putative plasma biomarker proteins. Anal. Bioanal. Chem. 404, 1089−1101.

Polanski, M., Anderson, N.L., 2006. A list of candidate cancer biomarkers for targeted proteomics. Biomark. Insights 1, 1−48.

Putnam, F.W., 1975. Alpha, beta, gamma, omega—the roster of the plasma proteins. Plasma Proteins 1, 57−130.

Razavi, M., Frick, L.E., Lamarr, W.A., Pope, M.E., Miller, C.A., Anderson, N.L., Pearson, T.W., 2012. High-throughput SISCAPA quantitation of peptides from human plasma digests by ultrafast, liquid chromatography-free mass spectrometry. J. Proteome Res. 11, 5642−5649.

Rein, B.J.D., Gupta, S., Dada, R., Safi, J., Michener, C., Agarwal, A., 2011. Potential markers for detection and monitoring of ovarian Cancer. J. Oncol. 2011, 475983.

Röst, H.L., Rosenberger, G., Navarro, P., Gillet, L., Miladinović, S.M., Schubert, O.T., Wolski, W., Collins, B.C., Malmström, J., Malmström, L., 2014. OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. Nat. Biotechnol. 32, 219−223.

Saenger, A., Beyrau, R., Braun, S., Cooray, R., Dolci, A., Freidank, H., Giannitsis, E., Gustafson, S., Handy, B., Katus, H., 2011. Multicenter analytical evaluation of a high-sensitivity troponin T assay. Clin. Chim. Acta 412, 748−754.

Schlautman, J.D., Rozek, W., Stetler, R., Mosley, R.L., Gendelman, H.E., Ciborowski, P., 2008. Multidimensional protein fractionation using ProteomeLab PF 2D for profiling amyotrophic lateral sclerosis immunity: a preliminary report. Proteome Sci. 6, 26.

Schoenherr, R.M., Saul, R.G., Whiteaker, J.R., Yan, P., Whiteley, G.R., Paulovich, A.G., 2015. Anti-peptide monoclonal antibodies generated for immuno-multiple reaction monitoring-mass spectrometry assays have a high probability of supporting Western blot and ELISA. Mol. Cell. Proteomics 14, 382−398.

Searle, B.C., 2010. Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies. Proteomics 10, 1265−1269.

Selevsek, N., Chang, C.-Y., Gillet, L.C., Navarro, P., Bernhardt, O.M., Reiter, L., Cheng, L.-Y., Vitek, O., Aebersold, R., 2015. Reproducible and consistent quantification of the *Saccharomyces cerevisiae* proteome by SWATH-mass spectrometry. Mol. Cell. Proteomics 14, 739−749.

Sehrawat, U., Pokhriyal, R., Gupta, A.K., Hariprasad, R., Khan, M.I., Gupta, D., Naru, J., Singh, S.B., Mohanty, A.K., Vanamail, P., Kumar, L., Kumar, S., Hariprasad, G., 2016. Comparative proteomic analysis of advanced ovarian cancer tissue to identify potential biomarkers of responders and nonresponders to first-line chemotherapy of carboplatin and paclitaxel. Biomark. Cancer 8, 43−56.

Services, U. S. D. O. H. A. H, 2015. In: Administration, F. A. D (Ed.), Analytical Procedures and Methods Validation for Drugs and Biologics.

Seshi, B., Raja, K., Chandramouli, K., 2011. Immobilized pH gradient-driven paper-based IEF: a new method for fractionating complex peptide mixtures before MS analysis. Clin. Proteomics 8, 10.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498−2504.

Shetlar, M., Bullock, J.A., Shetlar, C.L., Payne, R.W., 1955. Comparison of serum C-reactive protein, glycoprotein and seromucoid in cancer, arthritis, tuberculosis and pregnancy. Exp. Biol. Med. 88, 107−109.

Stamey, T.A., Yang, N., Hay, A.R., Mcneal, J.E., Freiha, F.S., Redwine, E., 1987. Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. N. Engl. J. Med. 317, 909−916.

Stinchcombe, T.E., Roder, J., Peterman, A.H., Grigorieva, J., Lee, C.B., Moore, D.T., Socinski, M.A., 2013. A retrospective analysis of VeriStrat status on outcome of a randomized phase II trial of first-line therapy with Gemcitabine, Erlotinib, or the combination in elderly patients (age 70 years or older) with stage IIIB/IV non−small-cell lung Cancer. J. Thorac. Oncol. 8, 443−451.

Takadate, T., Onogawa, T., Fukuda, T., Motoi, F., Suzuki, T., Fujii, K., Kihara, M., Mikami, S., Bando, Y., Maeda, S., Ishida, K., Minowa, T., Hanagata, N., Ohtsuka, H., Katayose, Y., Egawa, S., Nishimura, T., Unno, M., 2013. Novel prognostic protein markers of resectable pancreatic cancer identified by coupled shotgun and targeted proteomics using formalin-fixed paraffin-embedded tissues. Int. J. Cancer 132, 1368−1382.

Theis, J.D., Dasari, S., Vrana, J.A., Kurtin, P.J., Dogan, A., 2013. Shotgun-proteomics-based clinical testing for diagnosis and classification of amyloidosis. J. Mass Spectrom. 48, 1067−1077.

Tighe, P.J., Ryder, R.R., Todd, I., Fairclough, L.C., 2015. ELISA in the multiplex era: potentials and pitfalls. Proteomics Clin. Appl. 9, 406−422.

Ueland, F.R., Desimone, C.P., Seamon, L.G., Miller, R.A., Goodrich, S., Podzielinski, I., Sokoll, L., Smith, A., Van Nagell Jr., J.R., Zhang, Z., 2011. Effectiveness of a multivariate index assay in the preoperative assessment of ovarian tumors. Obstet. Gynecol. 117, 1289−1297.

Vaiopoulou, A., Gazouli, M., Papadopoulou, A., Anagnostopoulos, A.K., Karamanolis, G., Theodoropoulos, G.E., M'koma, A., Tsangaris, G.T., 2015. Serum protein profiling of adults and children with Crohn disease. J. Pediatr. Gastroenterol. Nutr. 60, 42−47.

Van Den Bergh, G., Arckens, L., 2004. Fluorescent two-dimensional difference gel electrophoresis unveils the potential of gel-based proteomics. Curr. Opin. Biotechnol. 15, 38−43.

Visintin, I., Feng, Z., Longton, G., Ward, D.C., Alvero, A.B., Lai, Y., Tenthorey, J., Leiser, A., Flores-Saaib, R., Yu, H., 2008. Diagnostic markers for early detection of ovarian cancer. Clin. Cancer Res. 14, 1065−1072.

Wang, L., Zhong, N., 2012. Application of the ProteomeLab™ PF2D protein fractionation system in proteomic analysis for human genetic diseases. Central Eur. J. Chem. 10, 836−843.

Wang, T.J., 2007. Significance of circulating troponins in heart failure if these walls could talk. Circulation 116, 1217−1220.

Wattiez, R., Falmagne, P., 2005. Proteomics of bronchoalveolar lavage fluid. J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 815, 169−178.

Wieser, A., Schneider, L., Jung, J., Schubert, S., 2012. MALDI-TOF MS in microbiological diagnostics—identification of microorganisms and beyond (mini review). Appl. Microbiol. Biotechnol. 93, 965−974.

Wolters, D.A., Washburn, M.P., Yates, J.R., 2001. An automated multidimensional protein identification technology for shotgun proteomics. Anal. Chem. 73, 5683−5690.

Wu, H.Y., Goan, Y.G., Chang, Y.H., Yang, Y.F., Chang, H.J., Cheng, P.N., Wu, C.C., Zgoda, V.G., Chen, Y.J., Liao, P.C., 2015. Qualification and verification of serological biomarker candidates for lung adenocarcinoma by targeted mass spectrometry. J. Proteome Res. 14, 3039−3050.

Wu, C., Duan, J., Liu, T., Smith, R.D., Qian, W.-J., 2016. Contributions of immunoaffinity chromatography to deep proteome profiling of human biofluids. J. Chromatogr. B 1021, 57−68.

Yocum, A.K., Chinnaiyan, A.M., 2009. Current affairs in quantitative targeted proteomics: multiple reaction monitoring-mass spectrometry. Brief. Funct. Genomics Proteomics 8, 145−157.

Yozwiak, N.L., Skewes-Cox, P., Stenglein, M.D., Balmaseda, A., Harris, E., Derisi, J.L., 2012. Virus identification in unknown tropical febrile illness cases using deep sequencing. PLoS Negl. Trop. Dis. 6, e1485.

Zhang, X., Belkina, N., Jacob, H.K., Maity, T., Biswas, R., Venugopalan, A., Shaw, P.G., Kim, M.S., Chaerkady, R., Pandey, A., Guha, U., 2015. Identifying novel targets of oncogenic EGF receptor signaling in lung cancer through global phosphoproteomics. Proteomics 15, 340−355.

Zhu, W., Smith, J.W., Huang, C.M., 2010. Mass spectrometry-based label-free quantitative proteomics. J. Biomed. Biotechnol. 2010, 840518.

Chapter 14

Molecular Cytogenetics in Molecular Diagnostics

H. Tönnies
Robert Koch-Institute, Berlin, Germany

14.1 INTRODUCTION

The fundamental cellular processes of DNA replication, DNA repair, mitosis, and meiosis ensure the high integrity of the human genome. However, mutations can occur in germline and somatic cells either induced or spontaneously. Depending on their physical size, these mutations manifest as altered sequences on the DNA level (e.g., gene variants) or gross alterations as numerical and structural chromosome aberrations. Chromosomal aberrations are the major cause of congenital anomalies, mental retardation, and infertility in humans. Furthermore, they account for approximately half of all early spontaneous pregnancy losses (Warburton, 2000). Overall, 1 in 120 live-born children is a carrier of a chromosomal abnormality. Half of these children are phenotypically conspicuous as a result of a chromosomal imbalance (Gardner et al., 2011). Four out of 1000 newborns are carriers of an unbalanced chromosome abnormality whereas 4.3 out of 1000 newborns show balanced chromosomal rearrangements without a phenotype. Partial losses (monosomy) or gains (trisomy) of euchromatic chromosomal material resulting in different clinical phenotypes have been described for all human chromosomes (Schinzel, 2001; http://www.ecaruca.net; https://decipher.sanger.ac.uk). In genetic prenatal, postnatal, and routine tumor genetic diagnostics, to date, conventional cytogenetics using classical karyotyping of chromosomes is the most widely used technique to characterize numerical and structural intrachromosomal and interchromosomal aberrations (Fig. 14.1).

In most countries, chromosome analysis is still the reference standard for the diagnosis and prognosis of congenital (inborn) and acquired (e.g., neoplasia) disorders. The precise characterization of aberrant cytogenetic findings is imperative for syndromologic assignment, phenotype–karyotype correlations, and genetic counseling in prenatal and postnatal diagnostics. In tumor genetics, conventional single-cell and metaphase cytogenetics is essential for disease monitoring, tumor staging, and research purposes to identify chromosomal regions harboring putative tumor suppressor and proto-oncogenes. However, limited chromosome-specific resolution obtained by conventional chromosome banding techniques makes the recognition and interpretation of masked or cryptic chromosome aberrations difficult if not impossible to ascertain.

Over the past 3 decades, molecular cytogenetic techniques based on fluorescence in situ hybridization (FISH) applications locating specific fluorescence-labeled nucleic acid sequences in interphase cells or metaphase chromosomes have become fast, sensitive, and important complementary tools in genetic diagnostics (Tönnies, 2002; Stumm and Tönnies, 2008). The use of diverse, locus-specific FISH probes and multicolor assays enhances the thorough characterization of numerical and complex chromosome aberrations regardless of their complexity, filling in parts of the gap between conventional chromosome banding analysis and molecular genetic studies on the DNA level. Until today, and even though next-generation high-throughput sequencing is becoming increasingly routine in diagnostic laboratories, multiple different molecular cytogenetic strategies and applications are used to generate routine diagnostic tests.

Numerous additional experimental-based techniques are used for comparative genomics and other fundamental research questions. This chapter focuses on diagnostic approaches used in routine laboratories to detect and characterize genetic changes beyond the base pair (bp)level. Compared with the first 2 decades of molecular cytogenetics, the development of new FISH-based approaches for

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00014-6
Copyright © 2017 Elsevier Ltd. All rights reserved.

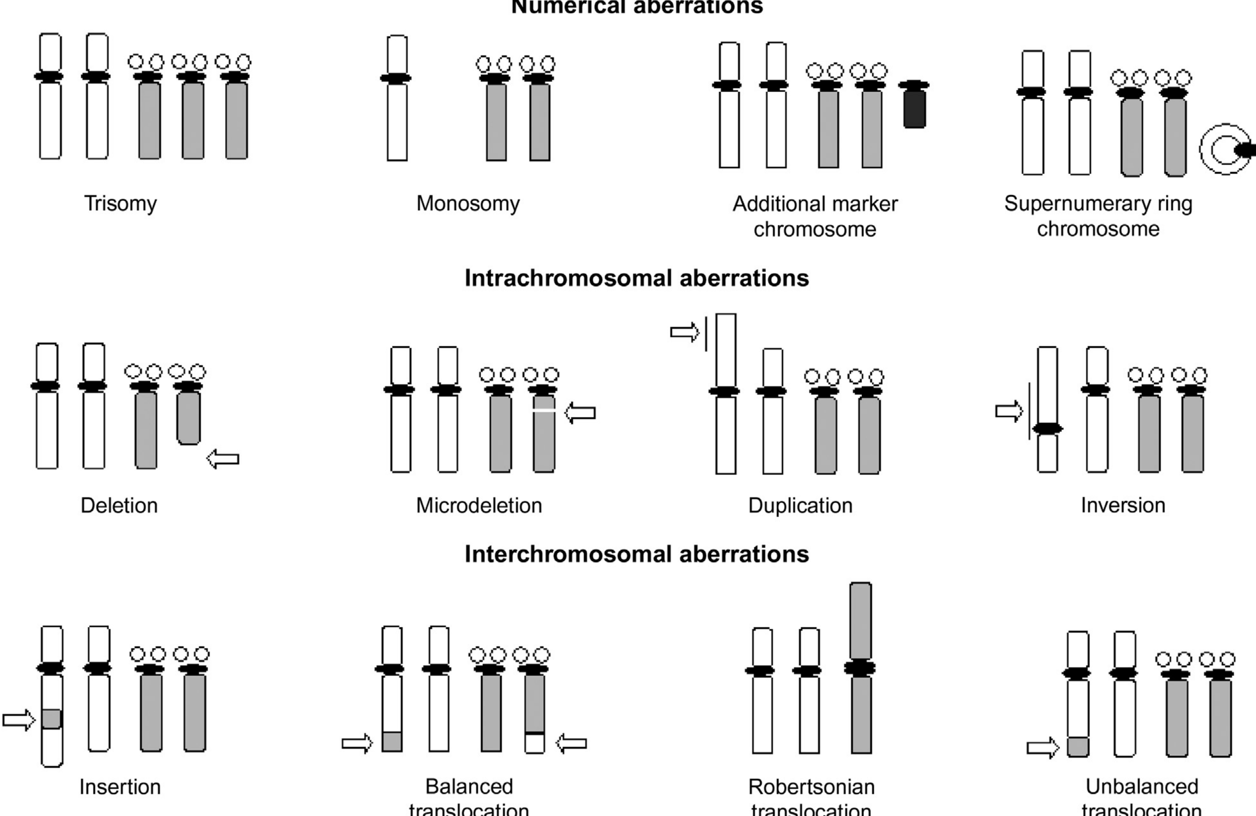


FIGURE 14.1 Schematic illustration of common numerical and structural intrachromosomal and interchromosomal aberrations in humans.

routine diagnostics has decreased because of the development of new high-resolution microarray-based approaches and high-throughput polymerase chain reaction (PCR)- or sequencing-based techniques, which have the potential to be automated. In addition they are fast and cheap, and are efficient compared with cell culture–dependent assays (for more details, see Chapters 8–10 and 20).

To understand array-based techniques and make professional decisions in a diagnostic context, it is essential to comprehend the development from conventional single-cell chromosome analysis to FISH-based approaches in the early 1980s and to modern array-based techniques. Furthermore, owing to the first array-based studies of the human genome, it became clear that single cell–based approaches such as interphase FISH (I-FISH) and karyotyping, together with quantitative PCR–based approaches, are essential for the validation and verification of whole-genome screening results to differentiate between pathogenic and benign copy number variations in the human genome. In the routine diagnostic setting, "indirect" genome-wide array methods and whole-exome/genome sequencing paved the way to investigate the exome/whole genome at high resolution. By contrast, focused, mainly single-cell ("direct") methods are cost-effective and indispensable for detecting specific submicroscopic alterations commonly found in clinically delineated syndromes (e.g., microdeletion syndromes) and neoplasias.

14.2 FROM CONVENTIONAL TO MOLECULAR CYTOGENETICS

The first drawings of human chromosomes were published more than 130 years ago (Arnold, 1879; Flemming, 1881; Hansemann, 1890; Cremer, 1985). It took several decades to develop stable protocols to prepare human chromosomes and establish the correct number of them. For clinical cytogeneticists, the 1960s were an exciting era. With the establishment of hypotonic treatment resulting in successful metaphase chromosome spreading (Hsu and Pomerat, 1953), Tijo and Levan (1956) managed to determine that the correct number of chromosomes in human diploid cells is 46. Three years later, Lejeune et al. (1959) published their observation that full trisomy 21 resulting in 47 chromosomes is the main cause of Down syndrome. The next technical improvement in cytology was the addition of colchicine to in vitro cultures for the accumulation of dividing cells before fixation (Moorhead et al., 1960).

In tumor genetics, the first finding of a chromosomal abnormality was the identification of a minute chromosome, named by its location of discovery, "Philadelphia chromosome," which was regularly found in peripheral blood cells of patients with chronic myeloid leukemia (Novell and Hungerford, 1960). In the late 1960s, with the development and adoption of new conventional cytogenetic banding techniques (Casperson et al., 1970) to detect numerical and structural chromosome abnormalities in humans, the era of

clinical cytogenetics began. In subsequent decades multiple additional conventional banding techniques were introduced for routine use (Verma and Babu, 1989; Barch et al., 1997) to detect chromosomal aberrations.

Today, the identification and characterization of numerical and structural chromosomal aberrations by karyotyping banded chromosomes are performed routinely in prenatal, postnatal, and tumor cytogenetic diagnostics as well as in basic genomic research.

Detectable chromosomal aberrations can be divided into numerical and structural aberrations (Fig. 14.1). In humans, numerical chromosome aberrations as aneusomies of whole chromosomes (e.g., trisomy 21, trisomy 18, or the chromosome constitution 45,X) are frequent, well-characterized chromosome abnormalities easily detectable by morphologic and numerical chromosome examination. Structural chromosomal abnormalities can be subgrouped into intrachromosomal (deletions, duplications, or inversions) and interchromosomal (balanced and unbalanced translocations or insertions) aberrations affecting more than one chromosome (Fig. 14.1; ISCN, 2013).

More laborious to identify by conventional cytogenetics alone are the so-called marker chromosomes. Marker chromosomes are structurally abnormal chromosomes, the origin of the euchromatic content of which cannot be determined by conventional cytogenetic analysis alone. Often supernumerary, their incidence varies from 0.3 to 3.7/1000 in newborns and patients with mental or developmental delay (Buckton et al., 1985; http://ssmc-tl.com/sSMC.html). Ring chromosomes, some also supernumerary (numerical aberration; e.g., owing to mitotic malsegregation), mainly originate from rearranged normal chromosomes as a result of deletions of the telomeric ends of the chromosome followed by ring formation (intrachromosomal aberration) (Tönnies et al., 2003a). However, the effectiveness of conventional cytogenetics directly depends on the size of chromosome aberrations appearing in human cells. The limited chromosome-specific banding resolution and assignment obtained by conventional chromosome banding makes the characterization and correct interpretation of complex and subtle chromosome aberrations difficult to ascertain and by nature is therefore often imprecise.

Furthermore, some chromosomal aberrations detected in an affected child are de novo, defined by a normal constitutional karyotype of the parents. Therefore, precise identification of the lost or duplicated chromosomal material is indispensable for any exact genotype–phenotype correlation as well as for diagnostic and possibly prognostic information. By conventional cytogenetic methods, rearrangements in the size of 5–10 megabases (Mb) affecting single chromosomes are detectable if high-resolution banding is achieved (approximately 500 bands/haploid genome). However, aberrations smaller than 5 Mb (e.g., microdeletion syndromes) (Table 14.1), complex chromosome aberrations involving three or more chromosomes, and marker chromosomes composed of unknown chromatic material often give unsatisfactory results using conventional cytogenetic techniques only. The dependence on dividing cells to prepare metaphase spreads, which is time-consuming, and the occurrence of chromosome aberrations not to be characterized exactly using classical banding techniques have been motives behind the development of alternative, more sensitive approaches based on the hybridization of DNA probes to patient DNA targets (e.g., metaphase chromosomes or interphase nuclei) to overcome the conventional resolution limitations inherent in classical cytogenetics.

14.3 FLUORESCENCE IN SITU HYBRIDIZATION

By in situ hybridization on metaphase spreads or interphase nuclei (e.g., nondividing cell populations), genetic changes can be analyzed at the single-cell level, allowing the simultaneous assessment of different chromosomes or chromosome regions and the determination of clonal variability or mosaicism (Fig. 14.2). It is a rapid, sensitive test for the detection of cryptic or subtle chromosomal changes allowing the identification of chromosomal alterations that are unresolved by conventional karyotyping. Classical in situ hybridization is based on the binding (hybridization or annealing) of complementary, single-stranded–labeled nucleic acids to the fixed and denatured target DNA of metaphase chromosomes, whole interphase nuclei, or DNA fibers. During hybridization, the probe penetrates to the target nucleic acid sequence and anneals to the complementary partner, resulting in DNA duplexes of the bound probe and the former single-stranded target. After removing unbound probe material by stringent washing, results can be inspected by microscopy.

First described by Pardue and Gall (1969) using radiolabeled, repetitive DNA probes hybridizing to mice cell preparations, the detection has been carried out by autoradiography enabling the morphologic (in situ) visualization of the presence of complementary nucleic acid sequences in the target material (John et al., 1969). Hybridization and detection of the first single-copy DNA sequences by autoradiography were described by Harper and Saunders (1981).

Molecular cytogenetics is based on FISH. Using nonradioactively labeled probes, Pinkel et al. (1986) introduced the basic protocol for FISH, improving the spatial resolution compared with radiolabeled probes. FISH is a sensitive and specific method that provides precise information about the physical location of a DNA sequence of interest in cell nuclei or chromosomes.

14.4 BASIC TECHNICAL ELEMENTS AND MATERIALS

FISH is a technique for the high-resolution visualization of labeled nucleic acid probes on target material (Fig. 14.2). It is a stepwise process beginning with the selection and

TABLE 14.1 Examples of Common Autosomal Microdeletion Syndromes in Clinical Diagnostics

Chromosomal Localization	Name	Involved Gene(s)/Hybridization Target	Online Mendelian Inheritance in Man[a] Entry
del(1)(p36.3)	Chromosome 1p36 deletion syndrome		#607872
del(4)(p16.3)	Wolf—Hirschhorn syndrome	Wolf—Hirschhorn critical region	#194190
del(5)(p15.2p15.3)	Cri-du-chat syndrome	Cri-du-chat critical region	#123450
del(7)(q11.23q11.23)	Williams—Beuren syndrome	*ELN*	#194050
del(8)(q24.1q24.1)	Trichorhinophalangeal syndrome/ Langer—Giedion syndrome	*EXT1, TRPS I, TRPS II*	#150230
del(11)(p13p13)	Wilms tumor, aniridia, genitourinary anomalies, mental retardation syndrome	*WT1, PAX6*	#194072
del(15)(q11q13)pat	Prader—Willi syndrome	*SNRPN*	#176270
del(15)(q11q13)mat	Angelman syndrome	*UBE3A*	#105830
del(16)(p13.3)	Rubinstein—Taybi syndrome	*CBP*	#180849
del(17)(p11.2p11.2)	Smith—Magenis syndrome	Smith—Magenis critical region	#182290
del(17)(p13.3)	Miller—Dieker syndrome	*LIS1*	#247200
del(20)(p11.23p11.23)	Alagille syndrome	*JAG1*	#118450
del(22)(q11.2q11.2)	Velocardiofacial/DiGeorge syndrome	*VCF*/DiGeorge critical region	#192430*188400

[a]*Online Mendelian Inheritance in Man at http://www.ncbi.nlm.nih.gov/omim.*

pretreatment of the target material, followed by the production and labeling of the appropriate probe to be hybridized, the hybridization process itself, and the detection, inspection, and documentation of the results. For the user, a huge number of variable targets, probes, and labeling procedures are available. The resolution of FISH analyses directly depends on two crucial elements: the target and the probe.

14.4.1 Targets for Fluorescence In Situ Hybridization

A wide variety of cellular materials can be used as targets for investigation by FISH. In routine diagnostics, in particular interphase nuclei and metaphase chromosomes of peripheral blood (routinely T-lymphocytes), primary fibroblasts, bone marrow cells, or, in prenatal diagnostics, amniocytes and chorionic villi cells are used after tissue cultivation (Fig. 14.2; Wegner, 1999). In leukemic specimens, in which the number of viable dividing cells is sometimes low, so-called hypermetaphase spreads obtained by long-term exposure of cells to colcemide can be produced (Seong et al., 1995). The yield of metaphase spread-like targets is higher compared with normal chromosome preparation with short-term colcemide treatment. Advantages compared with I-FISH are the ability to see a chromosome-like morphology and to minimize false-positive results owing to stochastic signal overlap, which is often seen in I-FISH experiments. Furthermore, without cultivation, buccal smear cells, sperm cells, and cells originating from the urinary tract can be used for I-FISH either to obtain primary information or to verify former results obtained by peripheral blood cell karyotyping or array—comparative genomic hybridization (CGH) analyses (see Section 15.10). In preimplantation genetic diagnosis (see Chapter 26), FISH for aneuploidy screening is routinely performed on polar bodies and, if allowed, on blastomeres (Delhanty et al., 1993; Fragouli, 2007).

Mainly used in basic research, e.g., for clone mapping, rather than in diagnostic units, FISH to DNA fibers (FIBER-FISH) using released chromatin fibers from interphase cells (Heng et al., 1992; Wiegant et al., 1992) is employed to study the structure and organization of mammalian genomes with high resolution. A great advantage of I-FISH on single cells is that it can also be performed on archived or nondividing material such as fresh-frozen and paraffin-embedded tissues slices and touch preparations of pathologic specimens. Resolution when

FIGURE 14.2 Flowchart summarizing the standard steps of fluorescence in situ hybridization. Target material (cell suspension including metaphase spreads and interphase nuclei) is dropped onto glass slides, pretreated, and denatured to prepare single-stranded DNA. After hybridizing the single-stranded, labeled probe to the target material and posthybridization wash, results can be analyzed by microscopic inspection. By using a whole-chromosome probe for chromosome 3 (*left*), a partial trisomy is visible. Using a locus-specific probe (e.g., YAC) (*right*), three signals are detectable in the interphase nucleus reflecting a partial trisomy 3.

using FISH depends directly on the target material. In the study of metaphase chromosomes, a resolution of 2–5 Mb can be attained, whereas in interphase nuclei the resolution varies between 2 Mb and 50 kb. When performing FIBER-FISH experiments, a resolution of 5–500 kb can be achieved (Speicher and Carter, 2005).

For reproducible hybridization results some targets need different pretreatment steps (e.g., proteolytic steps) for better penetration of the probe. For all target materials, a fixation step is required. In cytogenetics, conventional fixative (methanol/acetic acid, 3:1) is routinely used, but 70% ethanol fixation also provides good results. For FISH, formerly dividing cells are applied to glass slides for pretreatment by RNase or proteolytic digestion, if necessary. Before hybridization, the double-stranded target DNA has to be denatured by either chemical or heat treatment, "melting" the DNA double helices to single-stranded DNA (Schwarzacher and Heslop-Harrison, 2000).

14.4.2 DNA Probes for Fluorescence In Situ Hybridization

A variety of DNA probes generated by different amplification or cloning techniques are used for FISH. Probe DNA can be prepared by locus-specific PCR amplifying single genes, cloning of large human DNA fragments, chromosome microdissection, or flow-sorting of whole chromosomes. A large number of commercial FISH probes is available (Fig. 14.3). Region- or band-specific chromosomal DNA can be cloned in vectors of different sizes, such as fosmids, cosmids, plasmids, P1-derived artificial chromosomes, bacterial artificial chromosomes (BACs), or yeast artificial chromosomes (YACs). Yamada et al. (2011) introduced the visualization of fine-scale genomic structures using oligonucleotide libraries for high-resolution FISH. The choice of the vector system depends on the size of the DNA fragment to be cloned. Mapped and sequenced large insert clones are available for almost any chromosomal region and can easily be selected using Internet

databases (e.g., UCSC genome browser, Ensembl Cytoview, NCBI Map-Viewer).

14.4.3 Probe Labeling

Nucleic acid FISH probes are labeled and visualized either directly using fluorochrome-conjugated nucleotides (e.g., fluorescein isothiocyanate) or indirectly using reporter molecules (e.g., biotin-dUTP, digoxygenin-dUTP) (Schwarzacher and Heslop-Harrison, 2000). Nucleic acid—labeling approaches are based on the use of different enzymatic or chemical methods, and the choice of the labeling technique varies from laboratory to laboratory (Rautenstrauß and Liehr, 2002). Most labeling strategies are based on the enzymatic incorporation of label-conjugated nucleotides in a new synthesized DNA strand. Examples of these approaches are PCR (amplification and labeling of the probe) using degenerated oligonucleotide primers (Telenius et al., 1992), nick translation (no amplification but fragmentation of probe size) (Rigby et al., 1977), and random priming using hexamer primers resulting in probe amplification during the labeling procedure (Feinberg and Vogelstein, 1983). For small oligonucleotides, terminal fluorochrome-labeling reactions (Bauman et al., 1980) are commonly used.

Over the past decades, various other labeling approaches such as chemical cross-linking have been introduced (van Gijlswijk et al., 2001), and numerous kits are commercially available to label probe DNA with high efficiency.

For multicolor assays discriminating numerous targets simultaneously (for details, see Section 15.6), combinations of five fluorochromes can result in more than 24 colors using a ratio-labeling scheme (the same fluorochrome combination but different proportions) (Dauwerse et al., 1992; Nederlof et al., 1992), combinatorial labeling (a unique combination of different fluorochromes) as described by Nederlof et al. (1990), or the combined binary ratio labeling strategy introduced by Tanke et al. (1999).

A crucial step for all labeling systems is to achieve an optimal fragment size of the labeled FISH probe to ensure appropriate penetration and hybridization efficiency. For FISH experiments using locus-specific probes, the fragment size after labeling should not exceed 300 bp. Therefore, some labeling approaches also require fragmentation of probe DNA by DNase or ultrasonic treatment before in situ hybridization.

14.4.4 Hybridization, Posthybridization Wash, Detection, and Documentation

For DNA duplex formation, probe DNA and target DNA have to be single stranded (Fig. 14.2). For denaturation, the probe DNA dissolved in hybridization buffer (also known as hybridization mix) containing formamide and appropriate salt-concentrations is dissociated in a water bath at the appropriate temperature (depending on the probe composition). Many probes of higher complexity contain repetitive DNA elements that are scattered throughout the genome (e.g., whole-chromosome painting probes). To suppress their unspecific cross-hybridization, unlabeled competitor (Cot-1) DNA, a highly repetitive fraction of the human genome, is added to the hybridization mix. The Cot-1 DNA binds to the highly repetitive elements in the probe, suppressing nonspecific cross-hybridization to the target (chromosome in situ suppression hybridization) (Cremer et al., 1990; Hulten et al., 1991). After appropriate prehybridization or preannealing, the fluid hybridization mixture is applied to the target material for an appropriate time (hours to days), depending on the size and complexity of the probe and target. Nonspecific bound probe is washed off using stringent washes (low salt concentration) under thermal conditions that do not denature the target DNA duplexes (washing temperature lower than the melting temperature).

For direct-labeled probes, no detection steps are necessary, whereas indirect-labeled probes (e.g., biotin- or digoxigenin-labeled probes) have to be detected using an appropriate detection system. After counterstaining the target material using either 4,6-diamidino-2-phenylindole (DAPI) or propidium iodide, and application of an antifade solution to prevent bleaching of the fluorescent signals, hybridization results can be inspected visually using an epifluorescence microscope equipped with appropriate single- or multiband pass filter sets. Digital documentation of hybridization results is routinely done using a charge-coupled device (CCD) camera attached to a microscope, and connected to a computer with appropriate capturing and analysis software.

14.5 TYPES OF FLUORESCENCE IN SITU HYBRIDIZATION PROBES AND FLUORESCENCE IN SITU HYBRIDIZATION APPROACHES FOR METAPHASE AND INTERPHASE FLUORESCENCE IN SITU HYBRIDIZATION

The ability to detect and characterize chromosomal abnormalities in metaphase spreads and interphase cells using FISH has been greatly enhanced by the rapidly increasing availability of numerous chromosome- and locus-specific probes (Fig. 14.3 and Table 14.2). The choice of probe and the simultaneous use of multiprobe assays depend on the particular application in question.

14.5.1 Centromeric Satellite Probes

The first routinely used FISH probes were centromere-specific probes detecting highly repetitive centromeric

Whole chromosome probe
Satellite probe
Single fusion tumor probe
Chromosome arm specific probe
Microdeletion probe
Multicolor-banding probe
Subtelomeric probe

FIGURE 14.3 Schematic illustration of various fluorescence in situ hybridization probes used in routine cytogenetic diagnostics.

α-satellite DNA sequences (Cremer et al., 1986). In molecular diagnostics, these probes are mainly used for chromosome enumeration and marker chromosome identification (Figs. 14.3, 14.4A and B). Especially for interphase cell analysis, which can be performed on cytologic preparations as well as in sections of formalin-fixed and paraffin-embedded tissues, these probes are particularly attractive because of high sensitivity and excellent hybridization efficiencies. The hybridization time of these highly repetitive probes is short (only hours) and provides rapid results without the need for specific cytogenetic expertise for chromosome recognition or cell culture/metaphase preparation. So-called all-human centromeric probes, a mixture of all α-satellite repeats of the human genome hybridizing to all human centromeres simultaneously, are used to detect dicentric chromosomes or define neocentromeres and always lack α-satellite DNA sequences.

However, by using chromosome-specific centromeric probes for interphase cytogenetics, centromeric polymorphisms can result in split signals and therefore in false-positive findings. In addition, for some chromosomes, sequence homologies of the α-satellite sequences result in cross-hybridizations (e.g., chromosomes 13/21 and 14/22), which makes these probes less suitable for interphase cytogenetics. Furthermore, the use of centromeric probes for supernumerary marker chromosome detection gives no information about the euchromatic chromosome contents accounting for most phenotypic features.

14.5.2 Whole-Chromosome "Painting" Probes

Whole chromosome probes, also called painting probes, are DNA libraries representing a cocktail of DNA fragments of a single human chromosome (Deaven et al., 1986; Guan et al., 1993; Figs. 14.3, 14.4C and D). These DNA probes, obtained by chromosome flow sorting or chromosome microdissection (Fig. 14.5) (see Section 15.5.3), allow the labeling of individual chromosomes in metaphase spreads, and subsequently the identification and characterization of both numerical and interchromosomal structural aberrations as translocations and nonhomologous insertions. Partial chromosome probes (Fig. 14.3), generated by chromosome microdissection and representing the short or the long arm of chromosomes, are valuable tools for the detection of intrachromosomal pericentric inversions including the centromere.

In routine use, the application of appropriate whole-chromosome probes to ascertain chromosomal aberrations needs the prior knowledge (and therefore cytogenetic expertise) of the affected chromosome(s) in question. Otherwise, for example, for euchromatin-containing marker chromosome characterization, a whole repertoire of different DNA probes has to be hybridized and analyzed to narrow down or identify the origin of unknown chromosomal material (Blennow et al., 1995).

However, a locus-specific determination of the affected chromosomal region in intrachromosomal rearrangements, such as deletions and duplications, and paracentric inversions affecting only one chromosome arm, is not possible using whole or partial chromosome probes. The detection of small interchromosomal translocations involving only small regions of the chromosome ends often cannot be performed with adequate diagnostic sensitivity using whole-chromosome probes. In contrast to centromeric probes, chromosome enumeration by whole-chromosome probes can be performed only on metaphase spreads, which demands the availability of proliferating material.

14.5.3 Probe Generation by Chromosome Microdissection

Characterization of the composition and chromosomal origin of marker chromosomes or their parts can be performed straightforwardly by chromosome microdissection

TABLE 14.2 FISH Probes for the Characterization of Specific Chromosome Alterations

Probes for FISH	Hybridization Target	Patient Material	Aberration Type Detectable
Commercial Probes			
Centromere-specific α-satellite probes	Centromeres	Metaphase spreads, interphase cells	Numerical aberrations, identification of marker chromosome origin
Whole-chromosome probes and partial chromosome arm specific probes	Whole chromosomes	Metaphase spreads	Numerical aberrations, interchromosomal aberrations as balanced translocations or nonhomologous insertions
Locus-specific probes (microdeletion and microduplication probes)	Submicroscopic chromosomal loci	Metaphase spreads, interphase cells	Microdeletions and microduplications
Tumor probes (single-fusion probes, break-apart probes, double-fusion probes, oncogene amplification probes)	Submicroscopic fused chromosomal loci (translocation breakpoints), homogeneously staining regions (gene amplifications)	Interphase cells (metaphase spreads)	Chimeric gene fusions, gene amplifications, inversions, deletions
All telomeric probes	Telomere repeats	Metaphase spreads	Quantification of telomere size, loss of telomere
Subtelomeric probes	Subtelomeric regions	Metaphase spreads	Cryptic terminal deletions, duplications, translocations
Noncommercial Probes			
PCR-amplified microdissected or chromosomes or chromosomal bands (Micro-FISH)	Chromosomes or their parts	Patient and control metaphase spreads, forward and reverse "painting"	Marker chromosome characterization
PCR products in situ; PCR primer for specific chromosomal loci, genes, or alphoid sequences	Telomere repeats, microdeletions, single genes	Metaphase spreads	Telomeric aberrations, microdeletions, single gene deletions

FISH, fluorescence in situ hybridization; *PCR*, polymerase chain reaction.

(micro-FISH) (Meltzer et al., 1992). This method is based on the micromanipulated or laser capture–based chromosome dissection of the marker chromosome followed by PCR-mediated DNA amplification. By reverse (on normal metaphase spread) or forward (on patients' metaphase spreads) hybridization, the complete euchromatic content of the marker chromosome can be explored without using commercial probes (Fig. 14.5). A prerequisite for this assay is proliferating cell material from the patient. In addition to the fact that some chromosome aberrations such as inversions and small duplications are not detectable using dissected material as a probe, precise breakpoint identification is limited owing to hybridization on condensed metaphase chromosomes. Furthermore, there are practical limitations to micro-FISH for diagnostic purposes because specialized equipment (microdissection unit), technical skills, and profound cytogenetic knowledge regarding flawless recognition of the marker chromosome are necessary.

14.5.4 Region and Locus-Specific Probes

14.5.4.1 Locus-Specific Probes

To investigate submicroscopic chromosomal loci of the human genome, which are too small to be visualized by conventional cytogenetics, a wide spectrum of so-called locus-specific probes are used routinely in diagnostic FISH laboratories. Most of these probes are vector-cloned probes. For the molecular-cytogenetic verification of syndromes with submicroscopic gain or loss, such as the classical microdeletion and microduplication syndromes (Table 14.1), single-copy probes detecting the common region of interest are hybridized to patient cells (Fig. 14.4F). Although metaphase spreads are used routinely as targets for microdeletion syndromes, interphase nuclei are employed to detect microduplications because of the better separation of closely spaced hybridization signals (Shaffer and Lupski, 2000). Most commercial probes include a second control probe hybridizing to the

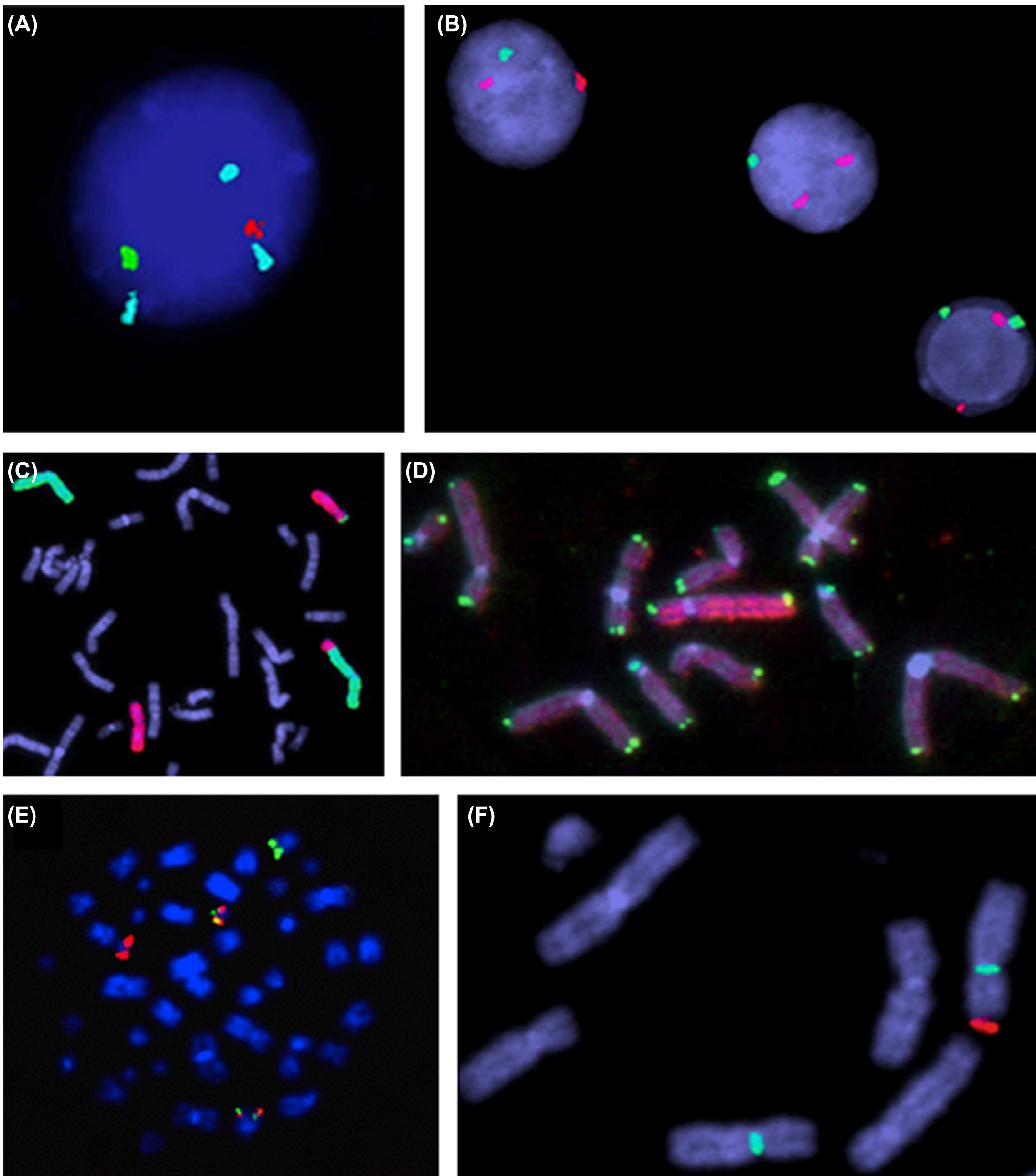


FIGURE 14.4 Sample combinations of single probes used in routine diagnostics. (A) Interphase nuclei of uncultured amniocytes hybridized with three different alphoid probes [cep-X (red (*light gray in print versions*)), cep-Y (green (*white in print versions*)), and cep-18 (blue (*gray in print versions*)). A normal male signal constitution has been detected for the chromosomes X and Y. The three blue (gray in print versions) signals gave evidence of trisomy 18 in the amniocytes. (B) Interphase—fluorescence in situ hybridization (FISH) using centromeric probes for chromosomes 7 (green (*white in print versions*)) and 9 (red (*light gray in print versions*)), detecting a monosomy 7 mosaicism. (C). Whole-chromosome probes for chromosome 3 (green (*white in print versions*)) and 11 (red (*light gray in print versions*)), detecting a reciprocal translocation t(3; 11). (D) Combinatorial hybridization of whole-chromosome painting for chromosome 4 (red (*light gray in print versions*)) and an all-telomere repeat probe. (E) Result of a double-fusion probe experiment performed on a bone marrow metaphase of a patient with chronic myelogenous leukemia patient detecting a Philadelphia-positive metaphase spread resulting in mixed-color signals (fusion signals) on the derivative chromosomes der(9) and der(22). (F) Microdeletion of the SHOX gene (red (*light gray in print versions*)) at the tip of one X-chromosome in a female. As a control probe, a cep-X (green (*white in print versions*)) has been hybridized. *(A) Courtesy of Markus Stumm, Berlin, Germany and (E) Courtesy of Ivan Loncarevic, Institute of Human Genetics and Anthropology, Jena, Germany.*

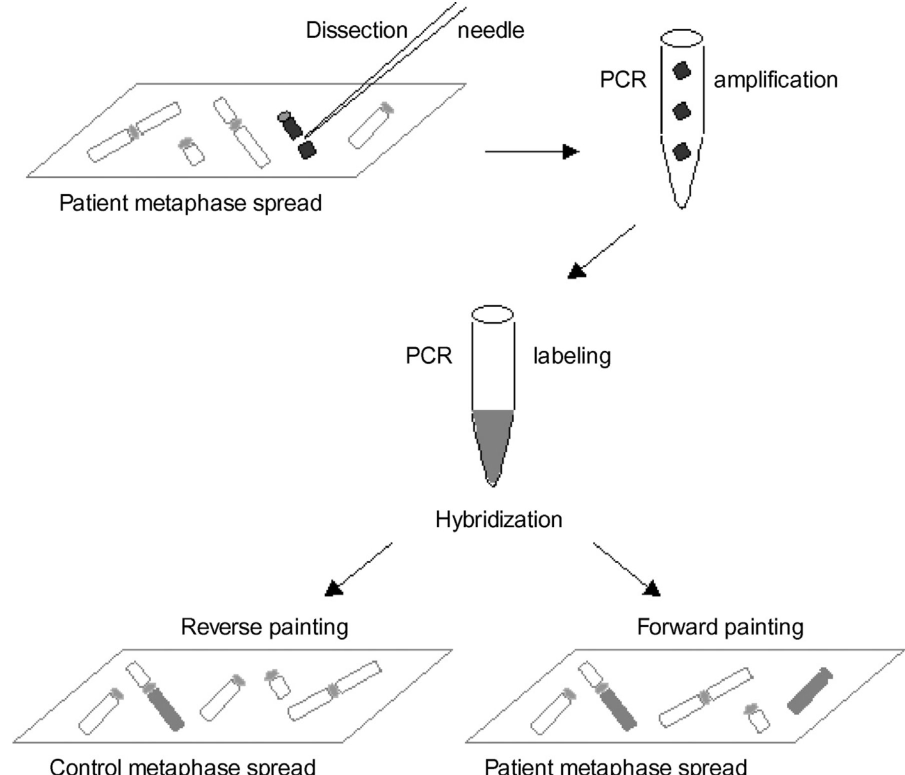


FIGURE 14.5 Flowchart summarizing the practical steps of marker chromosome microdissection fluorescence in situ hybridization. Using a glass dissection needle controlled by a micromanipulator, some copies of the marker chromosome are dissected and transferred to a microcentrifuge tube. Subsequently, chromosomal DNA is amplified and labeled by degenerate oligonucleotide primed−polymerase chain reaction (PCR). The resulting chromosome probe is hybridized to metaphases containing the marker (patient metaphase spread and forward painting) to verify regional authenticity of the probe. Simultaneously, the chromosomal composition of the marker is determined by hybridization of the probe to normal metaphase chromosomes.

unaffected chromosome arm to validate successful hybridization.

In prenatal diagnostics, differently labeled single probes can be used for rapid testing for chromosomal aneuploidies on uncultivated amniocytes (Kuo et al., 1991; Klinger et al., 1992; Ward et al., 1993; Stumm et al., 2006; Fig. 14.4A). Using this approach, a combination of locus-specific and alphoid probes of human chromosomes 21, 18, 13, X, and Y are hybridized in two combinations to detect the most common aneuploidies in humans. However, this FISH test is not designed to detect all possible chromosome aneuploidies; it needs amniotic fluid sampling and is therefore used as a fast and ancillary test for conventional cytogenetics. After the implementation of high-throughput next-generation sequencing, noninvasive prenatal testing is a straightforward alternative to circumvent amniocyte sampling.

14.5.4.2 Tumor Probes

In tumor genetics, most chromosomal rearrangements have major diagnostic and prognostic relevance. A number of single-copy locus-specific probes detecting characteristic translocations/fusion genes, inversions, deletions, and oncogene amplifications on the single-cell level are commercially available (Wang, 2002). These probes allow the investigation of metaphase spreads and interphase cells in one experiment. Furthermore, the detection of minimal residual disease and posttransplantation follow-up can be done with high sensitivity. Some of these locus-specific probes [locus-specific identifier (LSI)] locate translocation and inversion breakpoints precisely, reflecting chimeric gene fusions in neoplastic diseases, especially in leukemia (e.g., *BRC-ABL*). LSI probes can also be used to identify the origin of the amplified DNA that constitutes homogenously staining regions and double minutes that occur in a variety of tumor cells. Single-fusion dual-color FISH probes (Figs. 14.3 and 14.4E), which were the first available tumor probes, detect specific translocation breakpoints in interphase nuclei by the presence of nonrandomly distributed juxtaposed signals. However, in interphase diagnostics, these probes can have high rates of false-positive signals. Newer types of translocation probes in cancer genetics are break-apart probes detecting simple splits of normally proximate signals and so-called double-fusion probes detecting a fusion signal on both derivative chromosomes (Grand et al., 1998). Oncogene amplification probes, commonly used in solid tumor genetics, visualize gene amplifications in metaphase and interphase cells and clarify the identity of the genes involved.

14.5.4.3 Subtelomeric Probes and Telomere Repeat Probes

The highest gene concentrations in the human genome are in the subtelomeric regions of metaphase chromosomes (Saccone et al., 1992). These chromosomal regions are prone to rearrangements, which could give rise to cryptic aberrations probably accounting for 5−10% of unexplained moderate to severe cases of mental retardation, congenital anomaly, and spontaneous abortion (Knight et al., 1997, 1999; Anderlid et al., 2002; Yu et al., 2005). Subtelomeric probes were established for all chromosome ends (excluding the short arms of acrocentric chromosomes) to screen metaphase spreads for cryptic translocations or imbalances at the terminal euchromatic parts of the chromosomes.

The noncoding telomere repeat $(TTAGGG)_n$ distal to the subtelomeric region at the end of each eukaryotic chromosome protects the chromosome against rearrangements and fusion with other chromosomes (Fig. 14.4D). Using a repetitive all-telomere probe, ring chromosome formation can be pinpointed, especially in small supernumerary ring chromosomes.

Telomere shortening, which results from cell divisions over time, can lead to genomic instability and neoplasia (Shay et al., 1994). A quantitative FISH assay, using peptide nucleic acid (PNA) telomere oligonucleotide probes, resulting in stronger hybridization signals (with shorter probes, higher affinity, and specificity) compared with standard DNA oligonucleotide probes, has been described by Lansdorp et al. (1996). In addition, the PNA−DNA duplex has a higher melting temperature than the DNA−DNA duplex. Hybridizing these probes, the fluorescence intensity detected is directly proportional to the amount of telomere repeats.

14.5.5 Special Probe Types and Combinations

In the literature, several additional specific probe designs and combinations have been published over the years, but not all of these are used routinely in diagnostics. It would go beyond the scope of this chapter to list all of these in detail. All of the following FISH-based sensitive approaches detect a single or only a few chromosome alterations at a time.

14.5.5.1 Primed In Situ Labeling

Primed in situ labeling (PRINS) is a complementary FISH approach based on the in situ hybridization of short unlabeled DNA primers to metaphase spreads or interphase cells and subsequent in situ chain elongation catalyzed by a DNA polymerase (Koch et al., 1989, 1991). The product of this PCR in situ is visible owing to incorporation of labeled nucleotides. Depending on the primers used, this time- and cost-effective method allows the detection of centromeric α-satellite DNA (Hindkjaer et al., 1995), telomeric repeats (Krejci and Koch, 1998), microdeletions, and single-copy genes (Cinti et al., 1993; Kadandale et al., 2000; Tharapel et al., 2002). However, PRINS requires knowledge of the exact target sequence for the primers and high-quality target material.

14.5.5.2 Single-Copy Fluorescence In Situ Hybridization Probes

Rogan et al. (2001) introduced the use of a new generation of small single-copy FISH (sc-FISH) probes designed by computational sequence analysis of approximately 100-kb genomic sequences, bridging the gap between molecular genetic data and molecular cytogenetics. These labeled probes, produced by PCR, can be hybridized without preannealing or blocking, because sc-FISH probes lack repetitive DNA sequences. These short probes are produced directly from genomic DNA without recombinant DNA techniques. Applications of these probes include detection of microdeletion syndromes and submicroscopic deletions (Knoll and Rogan, 2003).

14.5.6 Morphology, Antibody, Chromosomes, In Situ Hybridization, and Fluorescence Immunophenotyping and Interphase Cytogenetics: Fluorescence Immunophenotyping and Interphase Fluorescence In Situ Hybridization

Morphology, antibody, chromosomes, in situ hybridization, and fluorescence immunophenotyping and interphase cytogenetics (FICTION) are methods combining the immunophenotyping of cells and FISH (Knuutila and Teerenhovi, 1989; Weber-Mathiesen et al., 1993). These approaches allow the examination of numerical and structural chromosome abnormalities of immunologically classified cells. They are used especially to detect tumor cells carrying chromosome abnormalities in mixed cell populations contaminated with normal cells as tissue sections or cytologic preparations. Martin-Subero et al. (2002) described the use of multicolor FICTION, allowing simultaneous detection of the morphologic and immunophenotypic characteristics of neoplastic cells together with the most frequent chromosomal aberrations.

14.6 MULTICOLOR FLUORESCENCE IN SITU HYBRIDIZATION SCREENING ASSAYS

For the simultaneous visualization of different chromosome aberrations in one experiment, a variety of multicolor FISH assays have been developed (Liehr et al., 2002; Tönnies, 2002). Some of the most important techniques for routine cytogenetics are described next (Table 14.3).

TABLE 14.3 Molecular Cytogenetic Multicolor Whole-Metaphase/Cell Scanning Techniques Used in Human Molecular Diagnostics

FISH Technique	Probe Setup	Hybridization Target (Patient Material)	Aberration Type Detectable	Specific Advantages of Technique	Diagnostic Limitations
Centromere-specific and centromeric- multiplex FISH	All centromere-specific α-satellite probes	Centromeres (metaphase spreads)	Numerical aberrations, identification of marker chromosome origin	Fast characterization of supernumerary marker chromosome origin, no need for specific cytogenetic expertise	No information about euchromatic content of marker
Telomere-multiplex-FISH, 12-color FISH	All subtelomeric regions	Subtelomeres (metaphase spreads)	Cryptic terminal deletions, duplications, and translocations	High sensitivity, simultaneous analysis of all chromosome ends involved in rearrangements	Small analysis spectrum restricted to chromosome ends, further analysis of affected chromosome(s) necessary
Spectral karyotyping/multi-fluor-FISH and technical modifications (color-changing karyotyping-FISH, combined binary ratio labeling-FISH, interspersed polymerase chain reaction multiplex FISH)	All human WCP-specific probes	Whole chromosomes (metaphase spreads)	Interchromosomal balanced and unbalanced translocations, euchromatic marker chromosome identification	Whole-metaphase scanning without prior probe selection, translocation partner detection in one experiment, fast characterization of euchromatic marker chromosome content	Insensitive detection of intrachromosomal aberrations such as deletions, duplications and inversions, imprecise breakpoint detection, need for nonoverlapping chromosomes
Chromosomal bar codes and cross-species color-segmenting FISH	yeast artificial chromosome clones, fragment hybrids, cross-species WCP mix	Whole chromosomes (metaphase spreads)	Interchromosomal and intrachromosomal aberrations such as translocations and gross deletions, duplications, inversions	Metaphase-wide detection of gross intrachromosomal and interchromosomal chromosome alterations and marker chromosomes	Suboptimal locus-specific resolution of breakpoints, need for nonoverlapping chromosomes
Multicolor-banding, multi-color banding	Microdissection-derived partial chromosome probes	Whole chromosomes (metaphase spreads)	Interchromosomal and intrachromosomal aberrations such as translocations, deletions, and inversions	Sensitive whole-karyotype detection chromosomal alterations with chromosome band—specific resolution	High costs and technically demanding, need for nonoverlapping chromosomes
Comparative genomic hybridization	Patient and control DNA	Whole control chromosomes (metaphase spreads)	Chromosomal interchromosomal and intrachromosomal imbalances, gene amplifications, marker chromosomes	Whole-genome scanning technique with no need for proliferating patient material, locus-specific detection of gene amplification, band-specific information on imbalance size	Insensitive detection of imbalances in subpopulation of cells, no detection of balanced aberrations, technically demanding
Array-based techniques (microarray)	Patient and control DNA	Spotted defined DNA probes (cloned sequences or oligonucleotides)	Chromosomal interchromosomal and intrachromosomal imbalances, gene amplifications, marker chromosomes	Whole-genome scanning technique with no need for proliferating patient material, high-resolution target specific detection of gene amplification, submicroscopic information on imbalances	Direct dependence on spotted targets, no detection of balanced translocations, missing of small mosaics

FISH, fluorescence in situ hybridization; *WCP*, whole-chromosome paint.
According to ISCN, 2013. An International System for Human Cytogenetic Nomenclature, (Shaffer, L.G., McGowan-Jordan, J., Schmid, M., eds.). S. Karger, Basel.

14.6.1 Centromere-Specific Multicolor Fluorescence In Situ Hybridization and Centromeric Multiplex Fluorescence In Situ Hybridization

Two all-human centromere-specific multicolor-FISH approaches for the subsequent determination of the exact origin of structurally abnormal, cytogenetically unidentifiable marker chromosomes were reported by Henegariu et al. (2001a) (centromeric multiplex FISH) and Nietzel et al. (2001) (centromere-specific multicolor FISH) (Fig. 14.6A). These one-step multicolor FISH assays allow the simultaneous characterization of all human centromeres using differently labeled centromeric satellite DNA as probes for all human chromosomes. However, by using centromeric α-satellite probes, only the centromeric parts of chromosomes can be explored, whereas other chromosome abnormalities affecting the euchromatin are excluded from investigation by using these assays.

Because most markers originate from the acrocentric chromosomes, Langer et al. (2001) developed an acroM-FISH assay using a probe mixture that consists of painting probes and centromere probes for chromosomes 13/21, 14/22, and 15, and a probe specific for ribosomal DNA, each labeled with a specific combination of fluorochromes. Using this assay, the origin of approximately 80% of all markers can be identified by one hybridization assay.

14.6.2 Multicolor Assays for Subtelomeric Rearrangements

Special subtelomeric multicolor probe mixes were established to scan metaphase spreads for subtle aberrations at the gene-rich ends of the chromosomes in a scanning assay (see Section 15.5.4.3). Henegariu et al. (2001b) reported the analysis of 41 chromosome ends simultaneously using a multicolor hybridization assay including DNA probes located near the end of these chromosomes (0.1 to 1 Mb from the telomere). In the same year, Brown et al. (2001) presented the 12-color FISH assay, which permits the screening of all telomeres in only two hybridizations. As an important parameter for successful hybridizations, all three techniques require well-spread metaphases (proliferating cells) with no cytoplasm.

14.7 MULTICOLOR WHOLE-METAPHASE SCANNING TECHNIQUES

In addition to the various locus-specific assays described in this chapter ("direct assays"), which often require the hybridization of a whole repertoire of probes for different chromosomes successively to narrow the overall composition of a chromosome aberration in single FISH experiments, some unbiased whole-metaphase or chromosome scanning approaches ("indirect assays") have been developed to characterize highly rearranged and unbalanced chromosome aberrations. Two major techniques, multi-fluor-FISH (M-FISH) (Speicher et al., 1996; Fig. 14.6B) and spectral karyotyping (SKY) (Schröck et al., 1996; Fig. 14.6C), were developed for the detection of nonhomologous structural and numerical chromosomal aberrations through the use of a 24-color whole-chromosome probe mix on individual metaphase chromosomes. Each homologous chromosome is displayed in a different color, based on the use of computer-generated false-color chromosome images and karyotyping. Although the principle of hybridizing whole-chromosome probes labeled by five different fluorochromes in different combinations (combinatorial labeling) is the same, both approaches employ a different technical strategy: spectral karyotyping combines Fourier spectroscopy, CCD imaging, and optical microscopy to measure at all points simultaneously in the sample emission spectra in the visible and near-infrared spectral range (Schröck et al., 1996). In contrast, M-FISH is based on classical epifluorescence microscopy using fluorochrome-specific optical filter sets for color discrimination (Speicher et al., 1996). That no prior knowledge of the affected chromosome(s) is required ("indirect analysis") counts as an advantage for both techniques. By both approaches, the detection of chromosome rearrangements such as balanced or unbalanced translocations and euchromatic marker chromosomes in a single hybridization is possible, even if the chromosome morphology is poor. Therefore, both methods are used routinely in clinical and tumor genetics (McNeil and Ried, 2000; Schröck and Padilla-Nash, 2000; Bayani and Squire, 2002; Teixeira, 2002).

However, using both techniques only the origin or constitution of an aberrant chromosome can be proved, whereas the assignment of an exact locus or breakpoint is not possible. A major drawback for routine use is poor sensitivity in detecting aberrations such as intrachromosomal deletions, duplications, and inversions. A large number of whole-chromosome probes labeled with different fluorochromes and the special technical equipment required, in particular for SKY analyses, restrict these scanning techniques to some specialized laboratories.

In 1999, Henegariu et al. presented a cost-effective alternative to SKY and M-FISH, called color-changing karyotyping (CCK). Compared to both techniques mentioned before, CCK uses only three fluorochromes to discriminate up to 41 DNA probes. Discrimination is done through the difference in signal strength between direct- and indirect-labeled chromosomes. Therefore, this approach can be used with a conventional three-filter fluorescent microscope located in every molecular-cytogenetic lab. To achieve a better breakpoint resolution

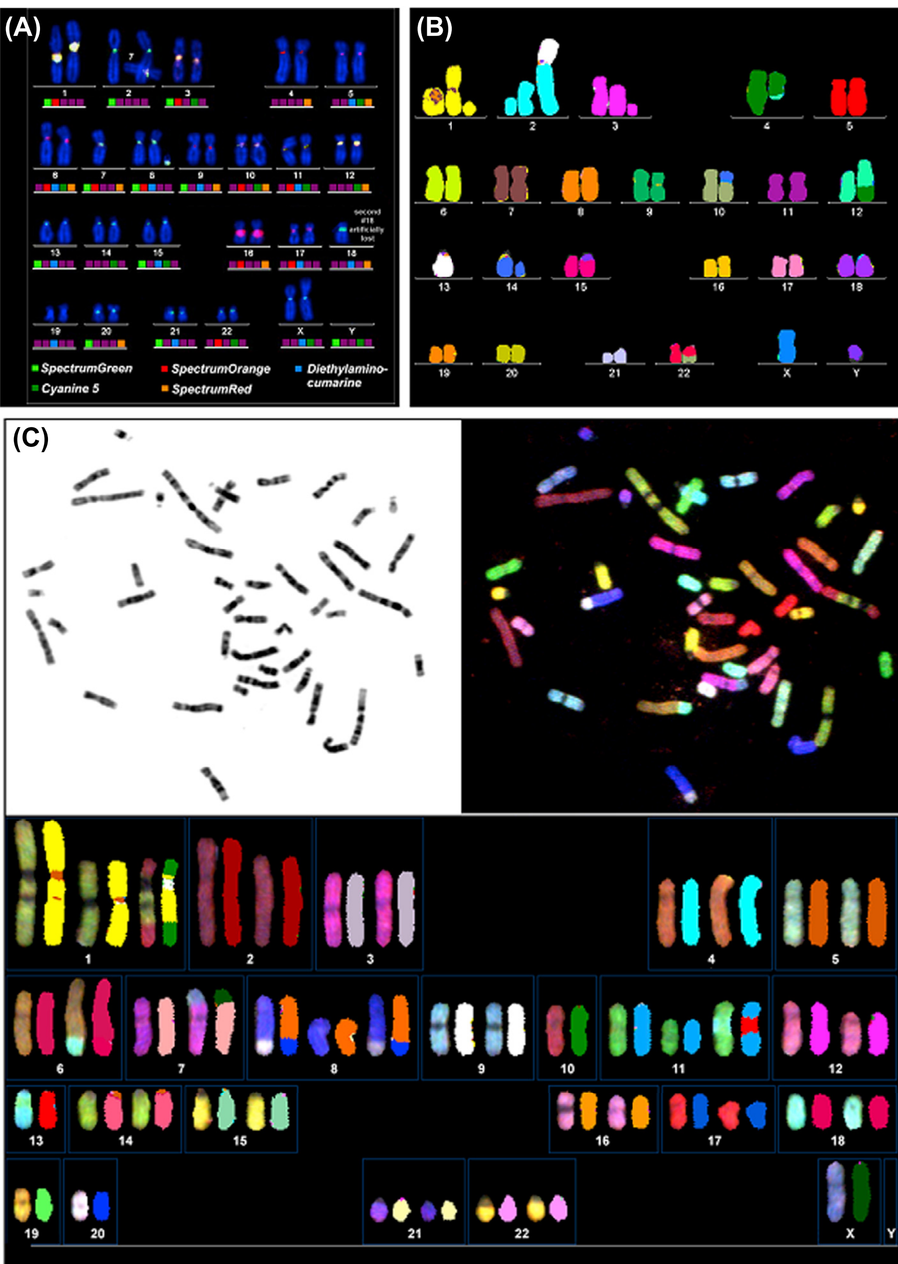


FIGURE 14.6 Different examples of multicolor fluorescence in situ hybridization (FISH) approaches. (A) A supernumerary marker chromosome could be characterized as a derivative chromosome 8 using the centromere-specific multicolor FISH approach. (B) Multi-fluor-FISH karyotype with complex chromosomal aberrations after radiation of the lymphocyte culture. (C). Spectral karyotyping analysis of a human multiple myeloma cell-line (NCI-H929, CRL-9068; ATCC, Manassas, VA). Several chromosomal aberrations were identified, e.g., translocations between chromosomes X and 7, 6 and 18, as well as 8 and 20, a dicentric chromosome 1 and a second dicentric chromosome 1 involved in a translocation or insertion with chromosome 10. *(A) Courtesy of Thomas Liehr, Institute of Human Genetics and Anthropology, Jena, Germany, (B) Courtesy of Christian Johannes, Essen, Germany. Hybridization; analysis carried out in collaboration with Metasystems, Germany), and (C) Courtesy of Isabell Grandy and Evelin Schröck, Institute of Clinical Genetics, Medical Faculty Carl Gustav Carus, Technical University, Dresden, Germany.*

of translocations, Aurich-Costa et al. (2001) developed a multicolor karyotyping technique based on the use of interspersed PCR (IRS-PCR)-generated whole-chromosome probes that display, in addition to color karyotyping, an R band—like hybridization pattern. This technique has been termed IRS-PCR multiplex FISH.

14.8 MULTICOLOR CHROMOSOME BANDING TECHNIQUES

Cytogenetics using black-and-white bar coding of chromosomes remains the most important and cost-effective standard method to identify chromosome aberrations.

However, because of the similar size and shade of some chromosome bands, detailed identification and chromosomal assignment of altered chromosome regions by conventional cytogenetic banding techniques alone is not always satisfactory. In addition, pronounced cytogenetic expertise is needed for aberration detection and characterization, which restricts the number of persons employing these methods. Over time, the idea came up to develop color bar coding of chromosomes, which allows the detection of interchromosomal and intrachromosomal aberrations as well as automated karyotyping of the colored chromosomes. The first bar-coding attempts, chromosomal bar codes (Lengauer et al., 1993), were obtained by FISH with pools of Alu-PCR products from yeast artificial chromosome clones containing human genomic DNA inserts.

Four years later, Müller et al. (1997) described their chromosome bar code for human chromosomes based on the application of a set of subregional DNA probes (based on human/rodent somatic cell hybrids) that distinguishes each chromosome in a single FISH assay. The result was a multicolor set of 110 distinct signals per haploid chromosome set. Using flow-sorted primate chromosomes of two gibbon species, Müller et al. (1998) also established the so-called cross-species color segmenting. With the exception of six chromosomes, all other human chromosomes can be differentiated by this hybridization assay in at least two and up to six segments. However, allowing the detection of inter and intrachromosomal aberrations, the resolution of classical banding patterns (approximately 400−600 bands/haploid genome) cannot be reached by these approaches (e.g., for breakpoint detection).

High-resolution multicolor banding (MCB) for refined FISH analysis of human chromosomes on the band and subband level was introduced by Chudoba et al. (1999) (Fig. 14.7A and B). This technique is based on changing fluorescence intensity ratios of overlapping DNA probes, labeled by five different fluorochromes. After computer-based assignment of distinctive pseudocolors to the overlapping and nonoverlapping hybridization signals along the chromosome, this approach allows a higher resolution than the other approaches and is independent of chromosome condensation. Using the MCB technique for single chromosomes, preferentially intrachromosomal aberrations such as deletions and pericentric or paracentric inversions can be characterized.

Mrasek et al. (2001) reported the use of human MCB probes for all chromosomes, permitting the straightforward characterization not only of intrachromosomal aberrations but also of interchromosomal translocations and breakpoint mapping (Fig. 14.7C). By using 138 microdissection-derived MCB probes, a resolution of 450 bands and more can be achieved, enabling high-resolution FISH banding for the detection of complex chromosome rearrangements (Liehr and Claussen, 2002). In 2008, Weise et al. presented the first BAC array-mapped genome-wide human MCB probe set including 169 region-specific microdissection libraries characterized in detail for their size and overlap.

14.9 WHOLE-GENOME SCANNING AND COMPARATIVE GENOMIC HYBRIDIZATION

All metaphase or chromosome scanning techniques mentioned before use metaphases of the patient to uncover chromosomal aberrations and, as a result, are experimentally demanding and labor-intensive. Moreover, resolution depends directly on the chromosomes used as targets. In tumor genetics, especially in solid tumor samples, the number of metaphases that can be prepared from these tissues is often small, if at all available. In addition, the chromosome morphology is often poor so that karyotyping is barely possible. Because numerous chromosomal abnormalities at a time are often expected, evaluation of all chromosomal imbalances can be time-consuming and costly.

To circumvent these problems, Kallioniemi et al. (1992) established a hybridization approach known as CGH, which relies on the use of genomic tumor DNA circumventing difficult cell culture and chromosome preparations of tumor tissue. CGH is a potent and reliable hybridization approach linking conventional cytogenetic and molecular genetic techniques. It allows a comprehensive analysis of the entire genome from tissue samples in just one experiment, providing global information about chromosomal imbalances and gene amplifications in the test genome (Tönnies et al., 2001a, 2003b; Tönnies and Lage, 2004; Fischer et al., 2004). In brief, CGH is based on the cohybridization of differentially labeled whole-genomic test (tumor) and normal (control or reference) DNA in a ratio of 1:1 to normal control (sex-matched) metaphase spreads (Fig. 14.8). Fluorochrome-labeled test and control DNA probes compete for hybridization on the target chromosomes. Because of deviations in fluorescence ratios of test DNA to control DNA, copy number changes cannot be detected by visual inspection alone (Fig. 14.9A). Therefore, software-assisted image capturing, karyotyping, and quantification of fluorescence intensities over the entire length of each chromosome are an integral part of this technique. The end point of CGH analysis is a so-called copy number karyotype displaying all chromosomal imbalances on a conventional cytogenetic resolution level (approximately 5−10 Mb) (Fig. 14.9B).

To enhance the sensitivity and resolution of CGH, Kirchhoff et al. (1998) reported on a high-resolution CGH approach based on modified CGH software. These authors applied standard reference intervals as detection criteria and

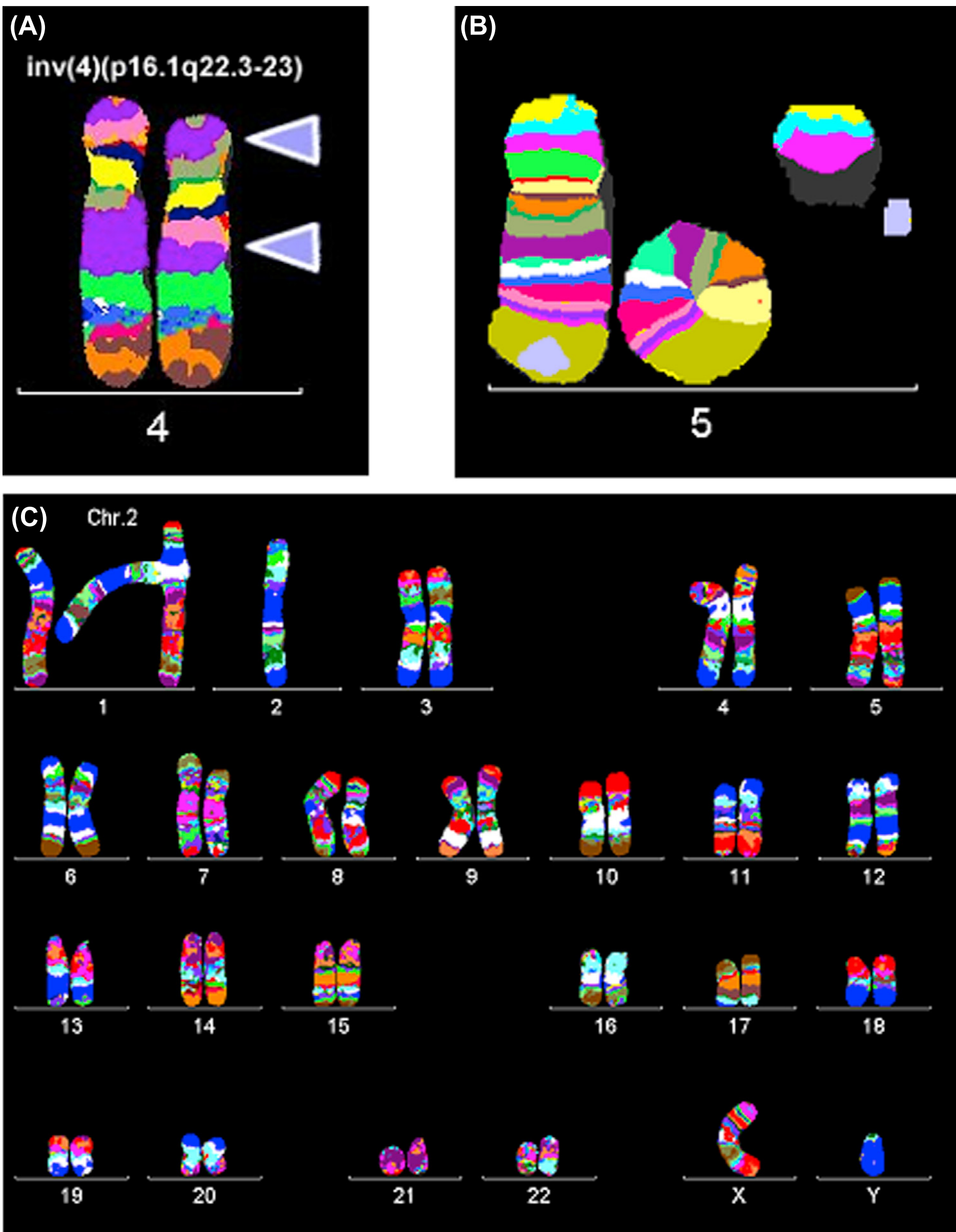


FIGURE 14.7 High-resolution multicolor banding (MCB) for refined fluorescence in situ hybridization (FISH) analysis of human chromosomes on the band and subband level. (A) MCB analysis showing a pericentric inversion (including the centromere) of chromosome 4. (B) MCB analysis of a ring chromosome 5 and a derivative chromosome 5 induced by radiation of lymphocyte cultures. (C) Example of the use of human MCB probes for all chromosomes simultaneously, permitting the straightforward characterization not only of intrachromosomal aberrations but also of interchromosomal translocations. By using 138 microdissection-derived MCB probes, a resolution of 450 bands and more can be achieved, enabling high-resolution FISH banding for the detection of complex chromosome rearrangements. *Pictures courtesy of Thomas Liehr and Christian Johannes, Institute of Human Genetics and Anthropology, Jena, Germany.*

detected deletions down to 3 Mb (Kirchhoff et al., 1999). CGH-based investigations provide information not only on the chromosomal assignment of a chromosomal imbalance but also on chromosomal band-specific origin. This approach allows DNA to be studied from any human source, even if the cells are not viable (e.g., paraffin-embedded tissue sections) or if the quantity of DNA is limited (see Section 15.5.3). Commonly employed in tumor genetics, a number of early reports describing prenatal and postnatal cytogenetic cases are available in the literature (Bryndorf et al., 1995; Levy et al., 1998; Tönnies et al., 2001a,b,c; 2003a,c). Even in preimplantation genetic diagnostics, the first experimental results were reported in 2002, revealing chromosomal imbalances in blastomeres or

FIGURE 14.8 Simplified flowchart summarizing the steps of comparative genomic hybridization (CGH) technique. Different-labeled control and test DNAs extracted from tissue cultures, whole blood, or paraffin-embedded material are hybridized under suppression conditions (Cot-1 DNA) on normal metaphase spreads. After image capturing and karyotyping of the metaphase chromosomes, fluorescence ratio profiles (quotient of test to control fluorescence intensity) for each chromosome are calculated. Mean profiles are plotted against the length of each chromosome. The center line (black) in the CGH profiles represents the balanced state of the chromosomal copy number (ratio 1.0). An upper threshold (green (*dark gray in print versions*)) is used to define a gain of chromosomal material [ENH (enhanced)], and a lower threshold (red (*gray in print versions*)) is used to interpret a loss of chromosomal material [DIM (diminished)]. CGH results are displayed in the so-called copy number karyotype.

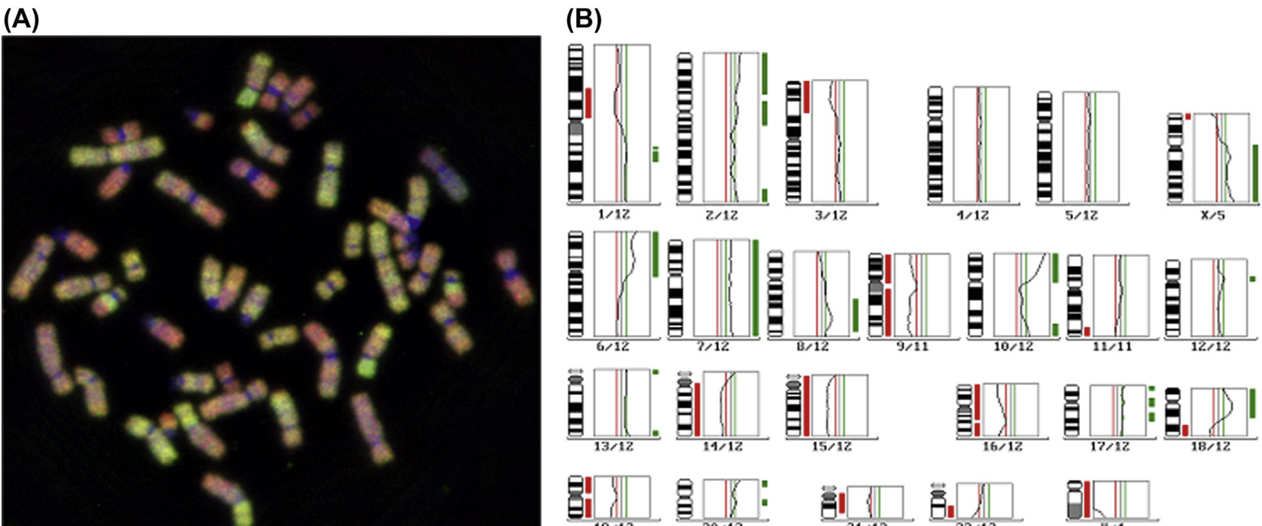


FIGURE 14.9 Comparative genomic hybridization (CGH) of DNA extracted from ovarian cancer tissue. (A) Metaphase spread hybridized with normal control DNA (red (*gray in print versions*)) and tumor DNA (green (*dark gray in print versions*)), and counterstained with 4,6-diamidino-2-phenylindole. A high number of fluorescence ratio differences are visible by inspection by eye. (B) Copy number karyotype of the same CGH analysis. Nearly all chromosomes are showing full (e.g., chromosome 7) or partial (e.g., chromosome 6) imbalances.

polar body cells using the DNA of single cells (Malmgren et al., 2002; Wells et al., 2002).

Nevertheless, by conventional CGH, the detection of balanced chromosome aberrations such as reciprocal translocations often found in liquid tumors (leukemias) is not possible. Furthermore, using whole-genomic DNA, information is restricted to the total amount of cells from which the DNA is extracted. Therefore, if only a small number of cells is affected by a chromosomal imbalance (heterogeneity in tumors or clonal aberrations), it will be missed as a result of contamination of the DNA with normal cells. However, the analysis of single cells is possible by upstream DNA amplification of single-cell DNA, as was shown by Klein et al. (1999, 2002). The most common method for DNA amplification is whole-genome amplification either by multiple displacement amplification or as PCR-based methods.

14.10 ARRAY-BASED TECHNIQUES (MICROARRAY)

Molecular cytogenetic FISH technologies, as described in this chapter, are based on the combination of molecular genetic and cytogenetic techniques detecting complementary nucleic acid sequences in situ on metaphase chromosomes or in the cell nuclei. However, the resolution level that is described for CGH as a genome-wide technique is restricted to approximately 5−10 Mb. To be independent of the need for high-quality metaphases and enhance the resolution of analyses, Solinas-Toldo et al. (1997) established a matrix-based CGH array that substitutes normal metaphase chromosomes as hybridization targets by well-defined genomic DNA fragments arrayed on a solid support, allowing automated analysis of genetic imbalances as small deletions or gene amplifications (Pinkel et al., 1998; Pollack et al., 1999; Fig. 14.10) (see Chapter 20). In 2001, Snijders et al. established a microarray using 2400 BAC inserts covering nearly the whole human genome to detect gains and losses in diploid, polyploid, and heterogeneous backgrounds. Veltman et al. (2002) described a DNA microarray spotted with a set of 77 human chromosome-specific subtelomeric probes as a target for array-based CGH. The dynamic range of the signal ratios obtained by these noncytogenetic assays is up to five times higher for matrix CGH compared with conventional CGH, allowing for better quantitative assessments of genetic imbalances (Wessendorf et al., 2002). Amplifications and deletions of small genomic regions not detectable by chromosome analyses or chromosomal CGH can be recognized with high resolution directly depending on the number of human DNA fragments spotted on the chip. By the use of 3500 inserts, a resolution of 1 Mb can easily be obtained (Coe et al., 2007). The robustness and simplicity of these

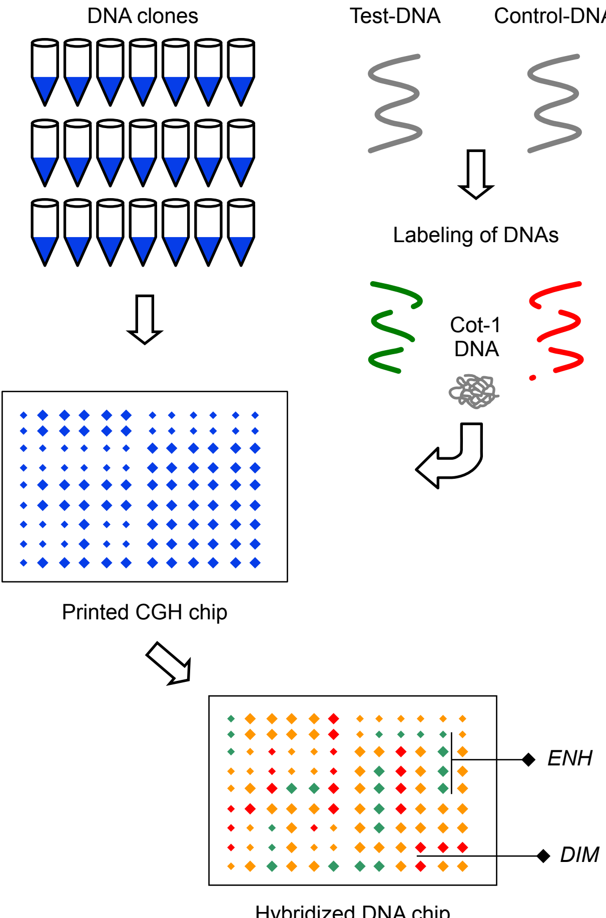


FIGURE 14.10 Schematic illustration of the matrix-comparative genomic hybridization (CGH) procedure. Cloned DNAs reflecting the human genome or their parts are printed on a CGH microarray and serve as a target for labeled control and test DNAs. After hybridization, the resulting relative fluorescence intensities are measured by computer-supported scanning. DNA loss in the test DNA results in red spots, and gain results in green spots. *Cot-1*, unlabeled competitor; *DIM*, diminished; *ENH*, enhanced.

microarrays make them highly suited for routine diagnostic use (Veltman, 2006; Vermeesch et al., 2007).

However, BAC inserts are typically large (80−200 kb), and therefore single-copy number aberrations smaller than 50 kb are not detectable with certainty. Shorter targets in higher numbers can be achieved by using oligonucleotides as targets (25−80 bp). Oligonucleotides synthesized directly onto the matrix offer the possibility of hybridizing on thousands of targets, depending on the company from which the array is ordered. Next-generation oligonucleotide arrays have up to millions of targets. Using custom-designed oligonucleotide arrays, almost nucleotide-level resolution can be achieved by synthesizing overlapping oligonucleotides with single-bp shifts. In 2010, Conrad et al. described the implementation of a 20−high-resolution oligonucleotide array carrying 42 million probes in total

and giving a median probe spacing of 56 bp across the entire genome.

Other platforms have been constructed initially for single-nucleotide polymorphism (SNP) studies. SNP arrays have been developed with 20 matched and mismatched probe pairs for each SNP allele. Hybridizations of test DNA with reduced complexity resulting from restriction enzyme digestion and adapter-ligation PCR are performed. SNP arrays make it possible to obtain an overview of chromosomal imbalances as well as copy number−neutral events such as loss of heterozygosity and uniparental disomies (UPD). The signal intensities of the matched probes are compared with in silico data sets. A 500-kb GeneChip shows a median spacing of 2.5 kb. Other arrays using 50-bp oligonucleotides are based on 650,000 different oligonucleotides with a median spacing of 2 kb. However, the targets of SNP arrays are not uniformly distributed across the genome compared with commercial oligonucleotide CGH arrays. The choice of platform depends on the purpose of the analysis. If small imbalances are sought, array CGH would probably be the best option. However, in cancer genetics, SNP arrays that also detected UPD may be the best platform. Some companies offer hybrid arrays combining the advantages of CGH and SNP arrays.

To detect a monogenic disorder in question and investigate the chromosomal complement [e.g., of an embryo in preimplantation genetic diagnosis (PGD)], "karyomapping," which was introduced in 2010 by Handyside et al., facilitates diagnostics at the single-cell level without prior patient- or disease-specific test development. This innovative approach combines monogenic disorder detection with high-resolution chromosome analysis in a single approach. In cases of PGD in which one parent is a carrier of a balanced reciprocal translocation, karyomapping is able to distinguish balanced and unbalanced embryos by identifying the parental haplotypes across the breakpoints.

14.11 CONCLUSIONS AND FUTURE PERSPECTIVES

Conventional and molecular cytogenetic approaches are important to characterize chromosomal aberrations deeply, and can distinguish between a simple gain or loss of genetic material and more complex aberrations such as insertions and cryptic abnormalities. Cytogenetic knowledge based on chromosomes and their parts has an important impact on genetic counseling and further clinical management. Therefore, both conventional and molecular cytogenetics serve important functions in modern medical diagnostics. Advances such as improved hybridization protocols and strategies, novel commercial and noncommercial probe sets, and hardware development such as image analysis software and specific filter sets for newly established fluorochromes facilitate the collection of new and important data for the understanding and diagnosis of genetic diseases.

More than a decade ago, genome-wide array techniques emerged that provided an opportunity to investigate the nuclear genome with high resolution in one experiment. These techniques allowed the identification of many new syndromes. Initially extremely expensive, this technique has increasingly replaced other (molecular) cytogenetic approaches because of its cost-effectiveness and wider acceptability. Directly transmitted imbalanced euchromatic variants without phenotypic effects have been detected by conventional cytogenetics and molecular cytogenetic methods for several years (Barber, 2005). Using high-resolution array platforms, a significant fraction of the genome has been identified to vary in copy number among apparently healthy individuals (Sebat et al., 2004; Iafrate et al., 2004; Redon et al., 2006). Current and near future challenges are to interpret these imbalances with caution, using several approaches including parental genotype, verification methods (mostly molecular cytogenetic methods or PCR-based technique), and last but not least, the patients' phenotype. Furthermore, one has to consider that a chromosomal abnormality may also be clinically significant even though it has been inherited from a healthy parent, because more complex mechanisms such as phenotypic variation, incomplete penetrance, imprinting, position effect, or even point mutation of a recessive gene may be involved. The amount of data mined today by different high-resolution molecular cytogenetic techniques requires a close interaction among the clinician, the genetic counselor, and last but not least, the molecular cytogeneticist using this palette of techniques. The main challenge for a molecular cytogenetic laboratory leader is to choose the best type of approach that is most useful and informative for the particular case in question. The identified genetic test might not be the newest or most "colorful" one. Sometimes, a black-and-white karyotype may be enough to detect the disease in question.

The introduction of next-generation sequencing technologies in routine diagnostics has resulted in an explosion of individual genomic information. The challenge is no longer the generation of data itself, but its correct interpretation. When exploring a higher number of genetic variants owing to the higher resolution of the technique that is used, there is always the higher chance of detecting unsolicited or secondary findings (Rigter et al., 2013). Before the advent of NGS, (molecular-) cytogenetic technologies existed with a low (targeted approaches) but significant chance (genome-wide analyses) of unsolicited findings (e.g., balanced chromosomal translocation). In modern molecular diagnostics, the main challenge is to generate a comprehensive analysis of an individual's genome in question and to limit possible unsolicited findings coming up with the approach in question (Botkin et al., 2015).

Over the past decades, molecular cytogenetics did not replace conventional cytogenetics in daily diagnostics; genome-wide microarray-based techniques (today, often a first-line diagnostic test in individuals with mental retardation) and high-resolution NGS are also not expected to replace (molecular) cytogenetics in routine (personalized) diagnostics. This palette of diagnostic techniques and applications currently available in routine genetic diagnostics is an indispensable tool for the individualized and responsible diagnosis of genetic diseases.

GLOSSARY

Aneuploidy A chromosome number that is not an exact multiple of the haploid chromosome number for a cell or organism. The irregular, aneuploid number of chromosomes or chromosomal regions can be higher or lower compared with the wild-type situation.

Centromere The constricted region of the chromosome where the sister chromatids are joined. For chromosome movement during mitosis and meiosis, the centromere is associated with spindle fibers.

Comparative genomic hybridization (CGH) Molecular cytogenetic whole-genome scanning technique based on the co-hybridization of labeled test and control DNA to normal metaphase spreads followed by the computer-assisted quantification of fluorescence intensities over the entire length of each chromosome resulting in a "copy number karyotype" reflecting chromosomal imbalances.

Fluorescence in situ hybridization (FISH) Molecular cytogenetic approach based on the duplex formation of complementary, single-stranded, and fluorescence labeled nucleic acid probes with target material (interphase cell and metaphase spread) allowing the exploration of the presence, number, and distinct location of genetic material in situ.

Hybridization Base pairing of complementary, single-stranded nucleic acids (DNA and RNA).

Interphase Any phase of the cell cycle between nuclear division other than mitosis (G1 phase, S phase, and G2 phase).

Metaphase The stage of nuclear division in the cell cycle in which highly condensed chromosomes are arranged in a plane between the two poles of the dividing cell. The condensed chromosomes of this phase are routinely used to identify chromosome aberrations.

Molecular cytogenetics Application of molecular biology techniques to cytogenetic preparations as metaphase spreads and interphase nuclei to identify chromosome abnormalities and RNA expression.

Mosaicism The presence of two or more genetically different cell populations in one organism.

Telomere Specialized DNA structure [(TTAGGG)n] capping the end of a chromosome involved in replication and stability of linear DNA molecules.

Translocation Interchange of parts of chromosomes after breakage and attachment to a different nonhomologous chromosome.

REFERENCES

Anderlid, B.M., Schoumans, J., Anneren, G., Sahlen, S., Kyllerman, M., Vujic, M., Hagberg, B., Blennow, E., Nordenskjold, M., 2002. Subtelomeric rearrangements detected in patients with idiopathic mental retardation. Am. J. Med. Genet. 107, 275−284.

Arnold, J., 1879. Beobachtungen über Kerntheilungen in den Zellen der Geschwülste. Virchows Arch. 78, 279−301.

Aurich-Costa, J., Vannier, A., Gregoire, E., Nowak, F., Cherif, D., 2001. IPM-FISH, a new M-FISH approach using IRS-PCR painting probes: application to the analysis of seven human prostate cell lines. Genes Chromosom. Cancer 30, 143−160.

Barber, J.C., 2005. Directly transmitted unbalanced chromosome abnormalities and euchromatic variants. J. Med. Genet. 42, 609−629.

Barch, M.J., Knutsen, T., Spurbeck, J.L., 1997. The AGT Cytogenetics Laboratory Manual. Lippincott-Raven, Philadelphia.

Bauman, J.G., Wiegant, J., Borst, P., van Duijn, P., 1980. A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochrome labelled RNA. Exp. Cell. Res. 128, 485−490.

Bayani, J.M., Squire, J.A., 2002. Applications of SKY in cancer cytogenetics. Cancer Invest. 20, 373−386.

Blennow, E., Nielsen, K.B., Telenius, H., Carter, N.P., Kristoffersson, U., Holmberg, E., Gillberg, C., Nordenskjold, M., 1995. Fifty probands with extra structurally abnormal chromosomes characterized by fluorescence in situ hybridization. Am. J. Med. Genet. 55, 85−94.

Botkin, J.R., Belmont, J.W., Berg, J.S., Berkman, B.E., Bombard, Y., Holm, I.A., Levy, H.P., Ormond, K.E., Saal, H.M., Spinner, N.B., Wilfond, B.S., McInerney, J.D., 2015. Points to consider: ethical, legal, and psychosocial implications of genetic testing in children and adolescents. Am. J. Hum. Genet. 97, 6−21.

Brown, J., Saracoglu, K., Uhrig, S., Speicher, M.R., Eils, R., Kearney, L., 2001. Subtelomeric chromosome rearrangements are detected using an innovative 12-color FISH assay (M-TEL). Nat. Med. 7, 497−501.

Bryndorf, T., Kirchhoff, M., Rose, H., Maahr, J., Gerdes, T., Karhu, R., Kallioniemi, A., Christensen, B., Lundsteen, C., Philip, J., 1995. Comparative genomic hybridization in clinical cytogenetics. Am. J. Hum. Genet. 57, 1211−1220.

Buckton, K.E., Spowart, G., Newton, M.S., Evans, H.J., 1985. Forty-four probands with an additional marker chromosome. Hum. Genet. 69, 353−370.

Casperson, T., Zech, L., Johhansson, C., 1970. Differential banding of alkylating fluorochromes in human chromosomes. Exp. Cell. Res. 60, 315−319.

Chudoba, I., Plesch, A., Lorch, T., Lemke, J., Claussen, U., Senger, G., 1999. High resolution multicolor-banding: a new technique for refined FISH analysis of human chromosomes. Cytogenet. Cell Genet. 84, 156−160.

Cinti, C., Santi, S., Maraldi, N.M., 1993. Localization of single copy gene by PRINS technique. Nucleic Acids Res. 21, 5799−5800.

Coe, B.P., Ylstra, B., Carvalho, B., Meijer, G.A., Macaulay, C., Lam, W.L., 2007. Resolving the resolution of array CGH. Genomics 89, 647−653.

Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., Fitzgerald, T., Hu, M., Ihm, C.H., Kristiansson, K., Macarthur, D.G., Macdonald, J.R., Onyiah, I., Pang, A.W., Robson, S., Stirrups, K., Valsesia, A., Walter, K., Wei, J., Wellcome Trust Case Control Consortium, Tyler-Smith, C., Carter, N.P., Lee, C., Scherer, S.W., Hurles, M.E., 2010. Origins and functional impact of copy number variation in the human genome. Nature 464, 704−712.

Cremer, T., 1985. Von der Zellenlehre zur Chromosomentheorie. Springer Verlag, Berlin, Heidelberg, New York, Tokyo.

Cremer, T., Landegent, J., Bruckner, A., Scholl, H.P., Schardin, M., Hager, H.D., Devilee, P., Pearson, P., van der Ploeg, M., 1986. 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.

Cremer, T., Popp, S., Emmerich, P., Lichter, P., Cremer, C., 1990. Rapid metaphase and interphase detection of radiation-induced chromosome aberrations in human lymphocytes by chromosomal suppression in situ hybridization. Cytometry 11, 110−118.

Dauwerse, J.G., Wiegant, J., Raap, A.K., Breuning, M.H., van Ommen, G.J., 1992. Multiple colors by fluorescence in situ hybridization using ratio-labelled DNA probes create a molecular karyotype. Hum. Mol. Genet. 1, 593−598.

Deaven, L.L., Van Dilla, M.A., Bartholdi, M.F., Carrano, A.V., Cram, L.S., Fuscoe, J.C., Gray, J.W., Hildebrand, C.E., Moyzis, R.K., Perlman, J., 1986. Construction of human chromosome-specific DNA libraries from flow-sorted chromosomes. Cold Spring Harb. Symp. Quant. Biol. 51, 159−167.

Delhanty, J.D., Griffin, D.K., Handyside, A.H., Harper, J., Atkinson, G.H., Pieters, M.H., Winston, R.M., 1993. Detection of aneuploidy and chromosomal mosaicism in human embryos during preimplantation sex determination by fluorescent in situ hybridisation, (FISH). Hum. Mol. Genet. 2, 1183−1185.

Feinberg, A.P., Vogelstein, B., 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132, 6−13.

Fischer, T.C., Gellrich, S., Muche, M., Sherev, T., Audring, H., Neitzel, H., Walden, P., Sterry, W., Tönnies, H., 2004. Genomic aberrations and survival in cutaneous T cell lymphomas. J. Invest. Dermatol. 122, 579−586.

Flemming, W., 1881. Beiträge zur Kenntnis der Zelle und ihrer Lebenserscheinungen. 3. Teil. Arch. Mikrosk. Anat. 20, 1−86.

Fragouli, E., 2007. Preimplantation genetic diagnosis: present and future. J. Assist. Reprod. Genet. 24, 201−207.

Gardner, R., Sutherland, G., Shaffer, L., 2011. Chromosome Abnormalities and Genetic Counseling, fourth ed. Oxford University Press.

Grand, F.H., Chase, A., Iqbal, S., Nguyen, D.X., Lewis, J.L., Marley, S.B., Davidson, R.J., Goldman, J.M., Gordon, M.Y., 1998. A two-color BCR-ABL probe that greatly reduces the false positive and false negative rates for fluorescence in situ hybridization in chronic myeloid leukemia. Genes Chromosom. Cancer 23, 109−115.

Guan, X.Y., Trent, J.M., Meltzer, P.S., 1993. Generation of band- specific painting probes from a single microdissected chromosome. Hum. Mol. Genet. 2, 1117−1121.

Handyside, A.H., Harton, G.L., Mariani, B., Thornhill, A.R., Affara, N., Shaw, M.A., Griffin, D.K., 2010. Karyomapping: a universal method for genome wide analysis of genetic disease based on mapping crossovers between parental haplotypes. J. Med. Genet. 47, 651−658.

Hansemann, D., 1890. Über assymmetrische Zellteilung in Epithelkrebsen und deren biologische Bedeutung. Arch. Pathol. Anat. 119, 299−326.

Harper, M.E., Saunders, G.F., 1981. Localization of single copy DNA sequences of G-banded human chromosomes by in situ hybridization. Chromosoma 83, 431−439.

Henegariu, O., Heerema, N.A., Bray-Ward, P., Ward, D.C., 1999. Colour-changing karyotyping: an alternative to MFISH/SKY. Nat. Genet. 23, 263−264.

Henegariu, O., Bray-Ward, P., Artan, S., Vance, G.H., Qumsyieh, M., Ward, D.C., 2001a. Small marker chromosome identification in metaphase and interphase using centromeric multiplex fish (CM-FISH). Lab. Invest. 81, 475−481.

Henegariu, O., Artan, S., Greally, J.M., Chen, X.N., Korenberg, J.R., Vance, G.H., Stubbs, L., Bray-Ward, P., Ward, D.C., 2001b. Cryptic translocation identification in human and mouse using several telomeric multiplex fish (TM-FISH) strategies. Lab. Invest. 81, 483−491.

Heng, H.H., Squire, J., Tsui, L.C., 1992. High-resolution mapping of mammalian genes by in situ hybridization to free chromatin. Hum. Mol. Genet. 1, 587−591.

Hindkjaer, J., Brandt, C.A., Koch, J., Lund, T.B., Kolvraa, S., Bolund, L., 1995. Simultaneous detection of centromere-specific probes and chromosome painting libraries by a combination of primed in situ labelling and chromosome painting (PRINS-painting). Chromosom. Res. 3, 41−44.

Hsu, T.C., Pomerat, C.M., 1953. Mammalian chromosomes in vitro. II. A method for spreading the chromosomes of cells in tissue culture. J. Hered. 44, 23−29.

Hulten, M.A., Gould, C.P., Goldman, A.S., Waters, J.J., 1991. Chromosome in situ suppression hybridisation in clinical cytogenetics. J. Med. Genet. 28, 577−582.

Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W., Lee, C., 2004. Detection of large-scale variation in the human genome. Nat. Genet. 36, 949−951.

ISCN, 2013. In: Shaffer, L.G., McGowan-Jordan, J., Schmid, M. (Eds.), An International System for Human Cytogenetic Nomenclature. S. Karger, Basel.

John, H., Birnstiel, M., Jones, K., 1969. RNA-DNA hybrids at the cytological level. Nature 223, 582−587.

Kadandale, J.S., Tunca, Y., Tharapel, A.T., 2000. Chromosomal localization of single copy genes SRY and SOX3 by primed in situ labeling (PRINS). Microb. Comp. Genomics 5, 71−74.

Kallioniemi, A., Kallioniemi, O.P., Sudar, D., Rutovitz, D., Gray, J.W., Waldman, F.M., Pinkel, D., 1992. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science 258, 818−820.

Kirchhoff, M., Gerdes, T., Rose, H., Maahr, J., Ottesen, A.M., Lundsteen, C., 1998. Detection of chromosomal gains and losses in comparative genomic hybridization analysis based on standard reference intervals. Cytometry 31, 163−173.

Kirchhoff, M., Gerdes, T., Maahr, J., Rose, H., Bentz, M., Dohner, H., Lundsteen, C., 1999. Deletions below 10 megabasepairs are detected in comparative genomic hybridization by standard reference intervals. Genes Chromosom. Cancer 25, 410−413.

Klein, C.A., Schmidt-Kittler, O., Schardt, J.A., Pantel, K., Speicher, M.R., Riethmüller, G., 1999. Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. Proc. Natl. Acad. Sci. U.S.A. 96, 4494−4499.

Klein, C.A., Seidl, S., Petat-Dutter, K., Offner, S., Geigl, J.B., Schmidt-Kittler, O., Wendler, N., Passlick, B., Huber, R.M., Schlimok, G., Baeuerle, P.A., Riethmüller, G., 2002. Combined transcriptome and genome analysis of single micrometastatic cells. Nat. Biotechnol. 20, 387−392.

Klinger, K., Landes, G., Shook, D., Harvey, R., Lopez, L., Locke, L., Osathanondh, R., Leverone, B., Houseal, T., Pavelka, K., Dackowski, W., 1992. Rapid detection of chromosome aneuploidies in uncultured amniocytes by using fluorescence in situ hybridization (FISH). Am. J. Hum. Genet. 51, 55–65.

Knight, S.J., Horsley, S.W., Regan, R., Lawrie, N.M., Maher, E.J., Cardy, D.L., Flint, J., Kearney, L., 1997. Development and clinical application of an innovative fluorescence in situ hybridization technique which detects submicroscopic rearrangements involving telomeres. Eur. J. Hum. Genet. 5, 1–8.

Knight, S.J., Regan, R., Nicod, A., Horsley, S.W., Kearney, L., Homfray, T., Winter, R.M., Bolton, P., Flint, J., 1999. Subtle chromosomal rearrangements in children with unexplained mental retardation. Lancet 354, 1677–1681.

Knoll, J.H., Rogan, P.K., 2003. Sequence-based, in situ detection of chromosomal abnormalities at high resolution. Am. J. Med. Genet. 121A, 245–257.

Knuutila, S., Teerenhovi, L., 1989. Immunophenotyping of aneuploid cells. Cancer Genet. Cytogenet. 41, 1–17.

Koch, J.E., Kolvraa, S., Petersen, K.B., Gregersen, N., Bolund, L., 1989. Oligonucleotide-priming methods for the chromosome- specific labelling of alpha satellite DNA in situ. Chromosoma 98, 259–265.

Koch, J., Hindkjaer, J., Mogensen, J., Kolvraa, S., Bolund, L., 1991. An improved method for chromosome-specific labeling of alpha satellite DNA in situ by using denatured doublestranded DNA probes as primers in a primed in situ labeling (PRINS). procedure. Genet. Anal. Tech. Appl. 8, 171–178.

Krejci, K., Koch, J., 1998. Improved detection and comparative sizing of human chromosomal telomeres in situ. Chromosoma 107, 198–203.

Kuo, W.L., Tenjin, H., Segraves, R., Pinkel, D., Golbus, M.S., Gray, J., 1991. Detection of aneuploidy involving chromosomes 13, 18, or 21, by fluorescence in situ hybridization (FISH) to interphase and metaphase amniocytes. Am. J. Hum. Genet. 49, 112–119.

Langer, S., Fauth, C., Rocchi, M., Murken, J., Speicher, M.R., 2001. AcroM fluorescent in situ hybridization analyses of marker chromosomes. Hum. Genet. 109, 152–158.

Lansdorp, P.M., Verwoerd, N.P., van de Rijke, F.M., Dragowska, V., Little, M.-T., Dirks, R.W., Raap, A.K., Tanke, H.J., 1996. Heterogeneity in telomere length of human chromosomes. Hum. Mol. Genet. 5, 685–691.

Lejeune, J., Turpin, R., Gautier, M., 1959. Chromosomic diagnosis of mongolism. Arch. Fr. Pediatr. 16, 962–963.

Lengauer, C., Speicher, M.R., Popp, S., Jauch, A., Taniwaki, M., Nagaraja, R., Riethman, H.C., Donis-Keller, H., D'Urso, M., Schlessinger, D., 1993. Chromosomal bar codes produced by multicolor fluorescence in situ hybridization with multiple YAC clones and whole chromosome painting probes. Hum. Mol. Genet. 2, 505–512.

Levy, B., Dunn, T.M., Kaffe, S., Kardon, N., Hirschhorn, K., 1998. Clinical applications of comparative genomic hybridization. Genet. Med. 1, 4–12.

Liehr, T., Heller, A., Starke, H., Rubtsov, N., Trifonov, V., Mrasek, K., Weise, A., Kuechler, A., Claussen, U., 2002. Microdissection based high resolution multicolor banding for all 24 human chromosomes. Int. J. Mol. Med. 9, 335–339.

Liehr, T., Claussen, U., 2002. Multicolor-FISH approaches for the characterization of human chromosomes in clinical genetics and tumor cytogenetics. Curr. Genom. 3, 213–235.

Malmgren, H., Sahlen, S., Inzunza, J., Aho, M., Rosenlund, B., Fridstrom, M., Hovatta, O., Ahrlund-Richter, L., Nordenskjold, M., Blennow, E., 2002. Single cell CGH analysis reveals a high degree of mosaicism in human embryos from patients with balanced structural chromosome aberrations. Mol. Hum. Reprod. 8, 502–510.

Martin-Subero, J.I., Chudoba, I., Harder, L., Gesk, S., Grote, W., Novo, F.J., Calasanz, M.J., Siebert, R., 2002. Multicolor-FICTION: expanding the possibilities of combined morphologic, immunophenotypic, and genetic single cell analyses. Am. J. Pathol. 161, 413–420.

McNeil, N., Ried, T., 2000. Novel molecular cytogenetic techniques for identifying complex chromosomal rearrangements: technology and applications in molecular medicine. Expert Rev. Mol. Med. 14, 1–14.

Meltzer, P.S., Guan, X.Y., Burgess, A., Trent, J.M., 1992. Rapid generation of region specific probes by chromosome microdissection and their application. Nat. Genet. 1, 24–28.

Moorhead, P.S., Nowell, P.C., Mellman, W.J., Battips, D.M., Hungerford, D.A., 1960. Chromosome preparations of leukocytes cultured from human peripheral blood. Exp. Cell. Res. 20, 613–616.

Mrasek, K., Heller, A., Rubtsov, N., Trifonov, V., Starke, H., Rocchi, M., Claussen, U., Liehr, T., 2001. Reconstruction of the female gorilla karyotype using 25-color FISH and multicolor banding (MCB). Cytogenet. Cell Genet. 93, 242–248.

Müller, S., Rocchi, M., Ferguson-Smith, M.A., Wienberg, J., 1997. Toward a multicolour chromosome bar code. Hum. Genet. 100, 271–278.

Müller, S., O'Brien, P.C., Ferguson-Smith, M.A., Wienberg, J., 1998. Cross-species colour segmenting: a novel tool in human karyotype analysis. Cytometry 33, 445–452.

Nederlof, P.M., van der Flier, S., Wiegant, J., Raap, A.K., Tanke, H.J., Ploem, J.S., van der Ploeg, M., 1990. Multiple fluorescence in situ hybridization. Cytometry 11, 126–131.

Nederlof, P.M., van der Flier, S., Vrolijk, J., Tanke, H.J., Raap, A.K., 1992. Fluorescence ratio measurements of double-labeled probes for multiple in situ hybridization by digital imaging microscopy. Cytometry 13, 839–845.

Nietzel, A., Rocchi, M., Starke, H., Heller, A., Fiedler, W., Wlodarska, I., Loncarevic, I.F., Beensen, V., Claussen, U., Liehr, T., 2001. A new multicolor-FISH approach for the characterization of marker chromosomes: centromere-specific multicolor-FISH (cenM-FISH). Hum. Genet. 108, 199–204.

Novell, P.C., Hungerford, D.A., 1960. A minute chromosome in human chronic granulocyte leukaemia. Science 132, 1497.

Pardue, M.L., Gall, J.G., 1969. Molecular hybridization of radioactive DNA to the DNA of cytological preparations. Proc. Natl. Acad. Sci. U.S.A. 64, 600–604.

Pinkel, D., Straume, T., Gray, J.W., 1986. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. Proc. Natl. Acad. Sci. U.S.A. 83, 2934–2938.

Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.L., Chen, C., Zhai, Y., Dairkee, S.H., Ljung, B.M., Gray, J.W., Albertson, D.G., 1998. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat. Genet. 20, 207–211.

Pollack, J.R., Perou, C.M., Alizadeh, A.A., Eisen, M.B., Pergamenschikov, A., Williams, C.F., Jeffrey, S.S., Botstein, D., Brown, P.O., 1999. Genome-wide analysis of DNA copy number changes using cDNA microarrays. Nat. Genet. 23, 41–46.

Rautenstrauß, B.W., Liehr, T., 2002. FISH Technology. Springer Lab Manual. Springer Verlag, Berlin, Heidelberg.

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., González, J.R., Gratacòs, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W., Hurles, M.E., 2006. Global variation in copy number in the human genome. Nature 444, 444−454.

Rigby, P.W., Dieckmann, M., Rhodes, C., Berg, P., 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with I. J. Mol. Biol. 113, 237−251.

Rigter, T., Henneman, L., Kristoffersson, U., Hall, A., Yntema, H.G., Borry, P., Tönnies, H., Waisfisz, Q., Elting, M.W., Dondorp, W.J., Cornel, M.C., 2013. Reflecting on earlier experiences with unsolicited findings: points to consider for next-generation sequencing and informed consent in diagnostics. Hum. Mutat. 34, 1322−1328.

Rogan, P.K., Cazcarro, P.M., Knoll, J.H., 2001. Sequence-based design of single-copy genomic DNA probes for fluorescence in situ hybridization. Genome Res. 11, 1086−1094.

Saccone, S., De Sario, A., Della Valle, G., Bernardi, G., 1992. The highest gene concentrations in the human genome are in telomeric bands of metaphase chromosomes. Proc. Natl. Acad. Sci. U.S.A. 89, 4913−4917.

Schinzel, A., 2001. Catalogue of Unbalanced Chromosome Aberrations in Man, second ed. De Gruyter, Berlin, New York.

Schröck, E., du Manoir, S., Veldman, T., Schoell, B., Wienberg, J., Ferguson-Smith, M.A., Ning, Y., Ledbetter, D.H., Bar-Am, I., Soenksen, D., Garini, Y., Ried, T., 1996. Multicolor spectral karyotyping of human chromosomes. Science 273, 494−497.

Schröck, E., Padilla-Nash, H., 2000. Spectral karyotyping and multicolor fluorescence in situ hybridization reveal new tumor-specific chromosomal aberrations. Sem. Hematol. 37, 334−347.

Schwarzacher, T., Heslop-Harrison, P., 2000. Practical In Situ Hybridisation. BIOS Scientific Publishers Limited, Oxford.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Månér, S., Massa, H., Walker, M., Chi, M., Navin, N., Lucito, R., Healy, J., Hicks, J., Ye, K., Reiner, A., Gilliam, T.C., Trask, B., Patterson, N., Zetterberg, A., Wigler, M., 2004. Large-scale copy number polymorphism in the human genome. Science 305, 525−528.

Seong, D.C., Kantarjian, H.M., Ro, J.Y., Talpaz, M., Xu, J., Robinson, J.R., Deisseroth, A.B., Champlin, R.E., Siciliano, M.J., 1995. Hypermetaphase fluorescence in situ hybridization for quantitative monitoring of Philadelphia chromosome-positive cells in patients with chronic myelogenous leukemia during treatment. Blood 86, 2343−2349.

Shaffer, L.G., Lupski, J.R., 2000. Molecular mechanisms for constitutional chromosomal rearrangements in humans. Annu. Rev. Genet. 34, 297−329.

Shay, J.W., Werbin, H., Wright, W.E., 1994. Telomere shortening may contribute to aging and cancer: a perspective. Mol. Cell. Differ. 2, 1−21.

Snijders, A.M., Nowak, N., Segraves, R., Blackwood, S., Brown, N., Conroy, J., Hamilton, G., Hindle, A.K., Huey, B., Kimura, K., Law, S., Myambo, K., Palmer, J., Ylstra, B., Yue, J.P., Gray, J.W., Jain, A.N., Pinkel, D., Albertson, D.G., 2001. Assembly of microarrays for genome-wide measurement of DNA copy number. Nat. Genet. 29, 263−264.

Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Dohner, H., Cremer, T., Lichter, P., 1997. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. Genes Chromosom. Cancer 20, 399−407.

Speicher, M.R., Carter, N.P., 2005. The new cytogenetics: blurring the boundaries with molecular biology. Nat. Rev. Genet. 6, 782−792.

Speicher, M.R., Ballard, S.G., Ward, D., 1996. Karyotyping human chromosomes by combinatorial multi-fluor FISH. Nat. Genet. 12, 368−375.

Stumm, M., Tönnies, H., 2008. Fluorescence in situ hybridization techniques in medical diagnostics. Expert Opin. Med. Diagn. 12, 1381−1390.

Stumm, M., Wegner, R.D., Bloechle, M., Eckel, H., 2006. Interphase M-FISH applications using commercial probes in prenatal and PGD diagnostics. Cytogenet. Genome Res. 114, 296−301.

Tanke, H.J., Wiegant, J., van Gijlswijk, R.P., Bezrookove, V., Pattenier, H., Heetebrij, R.J., Talman, E.G., Raap, A.K., Vrolijk, J., 1999. New strategy for multi-colour fluorescence in situ hybridisation: COBRA: COmbined Binary RAtio labelling. Eur. J. Hum. Genet. 7, 2−11.

Teixeira, M.R., 2002. Combined classical and molecular cytogenetic analysis of cancer. Eur. J. Cancer 38, 1580−1584.

Telenius, H., Carter, N.P., Nordenskjold, M., Ponder, B.A., Yunnacliffe, A., 1992. Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. Genomics 13, 718−725.

Tharapel, A.T., Kadandale, J.S., Martens, P.R., Wachtel, S.S., Wilroy Jr., R.S., 2002. Prader Willi/Angelman and DiGeorge/velocardiofacial syndrome deletions: diagnosis by primed in situ labeling (PRINS). Am. J. Med. Genet. 107, 119−122.

Tijo, J.H., Levan, A., 1956. The chromosomes of man. Hereditas 42, 1−6.

Tönnies, H., Stumm, M., Wegner, R.D., Chudoba, I., Kalscheuer, V., Neitzel, H., 2001a. Comparative genomic based strategy for the analysis of different chromosome imbalances detected in conventional cytogenetic diagnostics. Cytogenet. Cell. Genet. 93, 188−194.

Tönnies, H., Stumm, M., Neumann, L., Volleth, M., Grumpelt, U., Musebeck, J., Annuss, G., Neitzel, H., 2001b. Two further cases of WHS with unbalanced de novo translocation t(4;8) characterised by CGH and FISH. J. Med. Genet. 38, E21.

Tönnies, H., Schulze, I., Hennies, H.C., Neumann, L.M., Keitzer, R., Neitzel, H., 2001c. De novo terminal deletion of chromosome 15q26.1 characterised by comparative genomic hybridisation and FISH with locus specific probes. J. Med. Genet. 38, 617−620.

Tönnies, H., 2002. Modern molecular cytogenetic techniques in genetic diagnostics. Trends Mol. Med. 8, 246−250.

Tönnies, H., Neumann, L.M., Gruneberg, B., Neitzel, H., 2003a. Characterization of a supernumerary ring chromosome 1 mosaicism in two cell systems by molecular cytogenetic techniques and review of the literature. Am. J. Med. Genet. 121A, 163−167.

Tönnies, H., Huber, S., Kuhl, J.S., Gerlach, A., Ebell, W., Neitzel, H., 2003b. Clonal chromosome aberrations in bone marrow cells of Fanconi anemia patients: gains of the chromosomal segment 3q26q29 as an adverse risk factor. Blood 101, 3872−3874.

Tönnies, H., Hennies, H.-C., Spohr, H.L., Neitzel, H., 2003c. Characterization of the first supernumerary tricentric ring chromosome 1

mosaicism by conventional and molecular cytogenetic techniques. Cytogenet. Genome Res. 103, 28–33.

Tönnies, H., Lage, H., 2004. Chromosomal imbalances associated with drug resistance and thermoresistance in human pancreatic carcinoma cells. Eur. J. Cell Biol. 83, 591–601.

van Gijlswijk, R.P., Talman, E.G., Janssen, P.J., Snoeijers, S.S., Killian, J., Tanke, H.J., Heetebrij, R.J., 2001. Universal Linkage System: versatile nucleic acid labeling technique. Expert Rev. Mol. Diagn. 1, 81–91.

Veltman, J.A., Schoenmakers, E.F., Eussen, B.H., Janssen, I., Merkx, G., van Cleef, B., van Ravenswaaij, C.M., Brunner, H.G., Smeets, D., van Kessel, A.G., 2002. High-throughput analysis of subtelomeric chromosome rearrangements by use of array-based comparative genomic hybridization. Am. J. Hum. Genet. 70, 1269–1276.

Veltman, J.A., 2006. Genomic microarrays in clinical diagnosis. Curr. Opin. Pediatr. 18, 598–603.

Verma, R.S., Babu, A. (Eds.), 1989. Human Chromosomes: Manual of Basic Techniques. Pergamon Press, New York.

Vermeesch, J.R., Fiegler, H., de Leeuw, N., Szuhai, K., Schoumans, J., Ciccone, R., Speleman, F., Rauch, A., Clayton-Smith, J., Van Ravenswaaij, C., Sanlaville, D., Patsalis, P.C., Firth, H., Devriendt, K., Zuffardi, O., 2007. Guidelines for molecular karyotyping in constitutional genetic diagnosis. Eur. J. Hum. Genet. 15, 1105–1114.

Wang, N., 2002. Methodologies in cancer cytogenetics and molecular cytogenetics. Am. J. Med. Genet. 115, 118–124.

Warburton, D., 2000. Cytogenetics of reproductive wastage: from conception to birth. In: Mark, H.F.L. (Ed.), Medical Cytogenetics. Marcel Dekker, New York, USA, pp. 213–246.

Ward, B.E., Gersen, S.L., Carelli, M.P., McGuire, N.M., Dackowski, W.R., Weinstein, M., Sandlin, C., Warren, R., Klinger, K.W., 1993. Rapid prenatal diagnosis of chromosomal aneuploidies by fluorescence in situ hybridization: clinical experience with 4500 specimens. Am. J. Hum. Genet. 52, 854–865.

Weber-Matthiesen, K., Pressl, S., Schlegelberger, B., Grote, W., 1993. Combined immunophenotyping and interphase cytogenetics on cryostat sections by the new FICTION method. Leukemia 7, 646–649.

Wegner, R.D., 1999. Diagnostic Cytogenetics. Springer Lab Manual. Springer Verlag, Berlin, Heidelberg.

Weise, A., Mrasek, K., Fickelscher, I., Claussen, U., Cheung, S.W., Cai, W.W., Liehr, T., Kosyakova, N., 2008. Molecular definition of high-resolution multicolor banding probes: first within the human DNA sequence anchored FISH banding probe set. J. Histochem. Cytochem 56, 487–493.

Wells, D., Escudero, T., Levy, B., Hirschhorn, K., Delhanty, J.D., Munne, S., 2002. First clinical application of comparative genomic hybridization and polar body testing for preimplantation genetic diagnosis of aneuploidy. Fertil. Steril. 78, 543–549.

Wessendorf, S., Fritz, B., Wrobel, G., Nessling, M., Lampel, S., Goettel, D., Kuepper, M., Joos, S., Hopman, T., Kokocinski, F., Dohner, H., Bentz, M., Schwaenen, C., Lichter, P., 2002. Automated screening for genomic imbalances using matrixbased comparative genomic hybridization. Lab. Invest. 82, 47–60.

Wiegant, J., Kalle, W., Mullenders, L., Brookes, S., Hoovers, J.M., Dauwerse, J.G., van Ommen, G.J., Raap, A.K., 1992. High-resolution in situ hybridization using DNA halo preparations. Hum. Mol. Genet. 1, 587–591.

Yamada, N.A., Rector, L.S., Tsang, P., Carr, E., Scheffer, A., Sederberg, M.C., Aston, M.E., Ach, R.A., Tsalenko, A., Sampas, N., Peter, B., Bruhn, L., Brothman, A.R., 2011. Visualization of fine-scale genomic structure by oligonucleotide-based high-resolution FISH. Cytogenet. Genome Res. 132, 248–254.

Yu, S., Baker, E., Hinton, L., Eyre, H.J., Waters, W., Higgins, S., Sutherland, G.R., Haan, E., 2005. Frequency of truly cryptic subtelomere abnormalities – a study of 534 patients and literature review. Clin. Genet. 68, 436–441.

Chapter 15

Cytogenomics of Solid Tumors by Next-Generation Sequencing: A Clinical Perspective

D. Sie, E. Sistermans, D. Heideman and B. Ylstra
VU University Medical Center, Amsterdam, The Netherlands

15.1 A TUMOR PRESENTS

When a patient presents with symptoms of cancer, there is an immediate need for a proper diagnosis: "Do I indeed have a cancer?" And if it is indeed a tumor, the next questions are about consequence: "Is the tumor benign or malignant? Is it curable, or how long will I live? What are my treatment options? And how will all this affect the quality of my life and my loved ones?" To come to such a diagnosis, a stepwise process is undertaken that may require invasive diagnostic techniques in which a tissue biopsy is obtained. An essential step in this process is the pathologic review of the tumor material.

If the diagnosis points to a malignant tumor, the treating physician is generally joined by a multidisciplinary team, a so-called a "tumor board," to discuss treatment options. In The Netherlands the treating physician is frequently a medical oncologist or, depending on the type of tumor, a neuro-oncologist, dermatologist, pulmonologist, hematologist, pediatrician, or gynecologist. On the tumor board, surgeons and radiologists provide information about the anatomic localization and metabolic activity whereas the pathologist classically provides information about the macroscopic and microscopic appearance of the tissue and morphology of cell and nuclear structures as well as their (immuno)histopathologic makeup. For the treatment of cancer, surgery still provides one of the most common and adequate remedies. Drugs, like chemotherapy or targeted therapy, also form an important component of the treatment palette, which is frequently supported by radiology and/or immunology. The physical strength of the patient, as judged by the treating physician, is another important consideration that the board will take into account when formulating treatment options, because most cancer therapies are physically taxing. Advised by the treating physician, the patient ultimately decides which treatment option to select, considering the chances of successful cancer remission and the impact on the quality of life.

Apart from targeted mutation analysis by polymerase chain reaction (PCR), fluorescent in situ hybridization (FISH), and immunohistochemistry, the complete molecular genomic information of a tumor remains largely undisclosed to tumor boards. Considering that cancer is a disease of the genome (Hanahan and Weinberg, 2011), this molecular information is critical for an understanding of the etiology and vulnerabilities of cancer.

Important developments in molecular cytogenetic techniques have brought about improvements in cost-effectiveness, speed, accuracy, specificity, and sensitivity. The most notable developments are the increases in DNA sequencing speed, cost, and throughput through the introduction of next-generation sequencing (NGS). These developments will soon allow routine laboratories to sequence a genome for less than $1000 (Muir et al., 2016). Cancer genomes are sequenced with increased fidelity, resolution, and detail. As a consequence, an increasing number of molecular biomarkers are being discovered and validated. Another consequence is that existing markers for one (sub)type of cancer can be found that is applicable to alternative cancer types, which is made possible through the introduction of basket and umbrella trials (Hyman et al., 2015). Translation of these biomarkers into daily clinical practice requires the adoption of NGS techniques into routine diagnostics, which will have to accompany a

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00015-8
Copyright © 2017 Elsevier Ltd. All rights reserved.

transition in expertise in cytogenetic and molecular laboratories. The ultimate goal is to provide an increasingly comprehensive personalized diagnosis for all cancer patients by providing tumor boards with a complete molecular-cytogenomic picture of the tumors.

15.2 FROM MICROSCOPIC EXAMINATION TO MOLECULAR CYTOGENOMICS

For both cellular and molecular pathology, tumor tissue is obtained by cytology, needle biopsy, or surgical resection and is transported to the pathology laboratories. The tumor is examined macroscopically, dissected, and routinely fixed in formalin. Formalin fixation followed by paraffin embedding (FFPE) of tumor specimens has an advantage in that it can be stored at room temperature for extended periods while stably preserving the (intra)cellular histology of the tumor. Alternative fixation procedures, also geared to maintain tissue and cellular and nuclear structures, include formic acid and flash freezing in liquid nitrogen. Sections are made from the fixated material to perform chemical or immunohistochemical staining, followed by histopathologic diagnostics of the tumor tissue by microscopy. Various cytogenetic and molecular techniques have been developed that can make use of these same biosource specimens (Hastings et al., 2015). Storage of FFPE samples has led to the accumulation of decades' worth of patient materials. Much of these archive contents are accompanied by patients' history describing the treatment and disease course, and thereby serve as an invaluable resource for retrospective biomarker research (Blow, 2007; Scheinin et al., 2014). All projects investigating cancers, and particularly those with long survival, benefit from long-term clinical follow-up. For example, some patients with diffuse low-grade glioma can survive more than 25 years, and access to a large series of FFPE archival material was critical for us to learn that distal 10q deletion is significantly less common in tumors of these long-term survivors (Van Thuijl et al., 2014). We are fortunate in The Netherlands because studies with a large series of retrospective samples are facilitated through a nationwide pathology registry called PALGA (Casparie et al., 2007).

Besides long-term storage, a great advantage of FFPE is that the areas of tumor tissue can be marked by a pathologist and scraped away in a straightforward manner from normal cells before DNA isolation (Van Essen and Ylstra, 2012). However, the use of FFPE tumor specimens for molecular analysis also has drawbacks. Formalin fixation fixes all proteins in their position, which is perfect for morphology and immunohistochemistry, but DNA is damaged in the process. DNA derived from FFPE specimens is (1) damaged by nicks and gaps; (2) modified by the deamination—hydrolysis reaction of a cytosine into an uracil; (3) fragmented into small pieces; and (4) cross-linked, which limits the accessibility of the DNA in downstream analysis. RNA structures are similarly affected by the fixation process, and because of the unstable nature of messenger RNA (mRNA), its integrity can be heavily compromised ex vivo before formalin fixation. Furthermore, formalin fixation is detrimental for proteomics analysis.

Acquisition of the tissue material in itself also presents a number of challenges for cytogenomic diagnostics. Vital parts of the body, such as the brain, require a conservative resection strategy to obtain tissue material. Frequently only needle biopsies are taken that will have to suffice both histopathologic and molecular diagnostic procedures. This results in variable and often limited amounts of tissue material for DNA isolations. The tumor tissue material is frequently a heterogeneous mix of genetically distinct tumor cells with admixed normal cells, which makes downstream NGS data analysis all the more difficult. Even these dedicated data analysis approaches face limitations, because local sampling can cause specific tumor clones to be absent from the biopsy altogether. Once the tissue sample is collected, standardization and automation are not straightforward, particularly because pathology review before DNA or RNA isolation is crucial (Van Essen and Ylstra, 2012).

Despite these challenges, several molecular assays have been implemented successfully in clinical practice. Techniques such as the PCR, FISH, comparative genome hybridization (CGH), and array CGH assays (Ylstra et al., 2006; Hastings et al., 2015) have been implemented in the clinic through validation of the molecular procedures and meticulous description of analysis workflows in standard operating procedures. NGS has also made an entrance into cancer diagnostics; particularly for the analysis of smaller gene panels, most pathology laboratories in The Netherlands are up to speed (Sie et al., 2014; Boleij et al., 2015).

15.3 THE PROMISE OF LIQUID BIOPSY FOR CANCER DIAGNOSTICS

In hereditary genome diagnostics, blood has traditionally been the main source for DNA isolation and molecular assays. Acquisition of blood is highly standardized and uniform; it delivers a homogeneous and stable source for nucleic acid extraction. The use of robotics alleviates the manual workload while it increases the reproducibility and throughput of sample processing.

The liquid biopsy also holds great promise for cancer diagnostics. We anticipate liquid biopsy in cancer population screening, early detection, monitoring, and therapy selection. Apart from overcoming the limitations associated with tissue biopsies, liquid biopsy could offer solutions to the heterogeneity issue. Current assays with single molecular markers set a

precedent for the success of liquid biopsies in cancer diagnostics. Examples are the human papillomavirus test for cervical cancer (Schiffman et al., 2011), the prostate-specific antigen test for prostate cancer, and the fecal occult blood test for colorectal cancer screening.

In many academic hospitals, hereditary genome diagnostics has set the stage for liquid biopsies with the introduction of noninvasive prenatal testing (NIPT) by NGS (Tamminga et al., 2016). To perform NIPT, DNA is extracted from maternal blood, which includes a small fraction of fetal DNA. Subsequently, copy number NGS is performed to determine aneuploidy in minute amounts of fetal DNA (Straver et al., 2014). Diagnostic screens for cancer may strongly resemble the NIPT procedure. In fact, during a routine NIPT procedure, researchers at the hospital in Leuven (Belgium) incidentally observed genomic aberrations known to occur in Hodgkin lymphoma and were able to confirm an early stage of the disease in the patient (Amant et al., 2015). Given proper patient consent, NIPT thus has the potential to be used for early detection or even population screening of cancer.

There is currently no general consensus as to which liquid biopsy, method, or biomolecule is most informative for cancer diagnostics. This will depend on the cancer type and diagnostic goal, such that alternate liquid sources and analytical methods will be applicable. Existing tumor tissue biomarkers may be retrieved from liquid biosources such as cell-free DNA, which is shed by the tumor in the blood circulation. Alternatively, surrogate signatures for tumor tissue biomarkers may be obtained from mRNA, that is, extracted from blood platelets, as we demonstrated at the Cancer Center Amsterdam (Best et al., 2015). Besides blood as a source for liquid biopsy, other bodily fluids are under investigation as a source for cancer screening and monitoring, including sputum, urine, cerebrospinal fluid, cervical, anal, penile, and oral swabs. The primary challenge for the clinical implementation of liquid biopsy diagnostics is, of course, the sensitivity of the DNA test. NGS along with novel PCR techniques such as droplet digital PCR currently offers the highest sensitivity to detect mutations and aberrations. DNA remnants can be limited in the various liquid biopsies, and hence a crucial step in the process is to enrich for circulating tumor DNA and keep the background of nonaberrant signal low. Besides these laboratory challenges, special bioinformatics procedures will have to be adopted to allow the introduction of liquid biopsy cancer cytogenomics into clinical diagnostic practice.

15.4 CYTOGENOMIC APPLICATIONS

15.4.1 Chromosomal Aberrations

In cancer research, we use copy number profiling to assess prognostic and predictive biomarkers in various tumor types including colon, low-grade glioma, and head and neck tumors (Van Thuijl et al., 2014; Haan et al., 2014; Smeets et al., 2009). A routine diagnostic application in our clinic is the assessment of the clonal relationship among multiple tumors from a single patient. The assay determines whether multiple tumors can be regarded as independent tumors or a single primary tumor with metastasis. The clonal relationship can be interpreted in an automated fashion by calculating a likelihood ratio to distinguish tumor pairs (Ostrovnaya et al., 2010). In our daily practice, we complement this procedure by calculating a correlation of aberrant segment values (Fig. 15.1), which has brought valuable insights to the clinic for effective cancer treatment (Kuiper et al., 2015). Copy number aberrations are classically investigated by FISH, karyotyping of condensed chromosomal structures, CGH, and later, array CGH (Ylstra et al., 2006). It is a time-consuming and expensive undertaking to design probes for the FISH and CGH techniques and requires prior knowledge about the genomic sequences of interest. Whole-genome sequencing (WGS) offers an unbiased approach because the complete DNA content of the acquired nuclei is simply read out by NGS. An equivalent resolution to microarrays (180,000 features) can be achieved with a limited amount of data in which only a tenth of the (human) genome (0.1× coverage) is actually sequenced. This low-pass or shallow-WGS (sWGS) allows for cheap and efficient copy number profiling, even when using DNA isolated from FFPE tissue material (Scheinin et al., 2014). The general approach involves dividing the human reference genome into nonoverlapping fixed-size bins. After mapping the reads to the reference genome, the raw read count is determined by the number of mappings contained in each bin. After guanine-cytosine content and "mappability" correction, filtering, segmentation, and calling, this read count reflects an accurate copy number state of the analyzed sample.

15.4.2 Actionable Hot-Spot Gene Panel

Investigation of predictive biomarkers to assess tumor susceptibility to targeted therapies can use a focused approach by considering only a limited set of relevant genes to assay. For instance, patients with melanoma are tested for the BRAF p.Val600Glu mutation. Th presence of BRAF p.Val600Glu indicates a susceptibility to vemurafenib, an enzyme inhibitor that acts on the BRAF/MEK/ERK pathway to cause programmed cell death in late-stage melanoma (Bollag et al., 2012). Besides BRAF there are currently around 50 other so-called "actionable" genes. Instead of targeting all exons of these genes, a "hot-spot" approach can be implemented to target, for example, only around 250 of the commonly mutated loci with known clinical relevance. Currently, the turnaround time from sample to "hot-spot" result is 1 week in our clinical practice

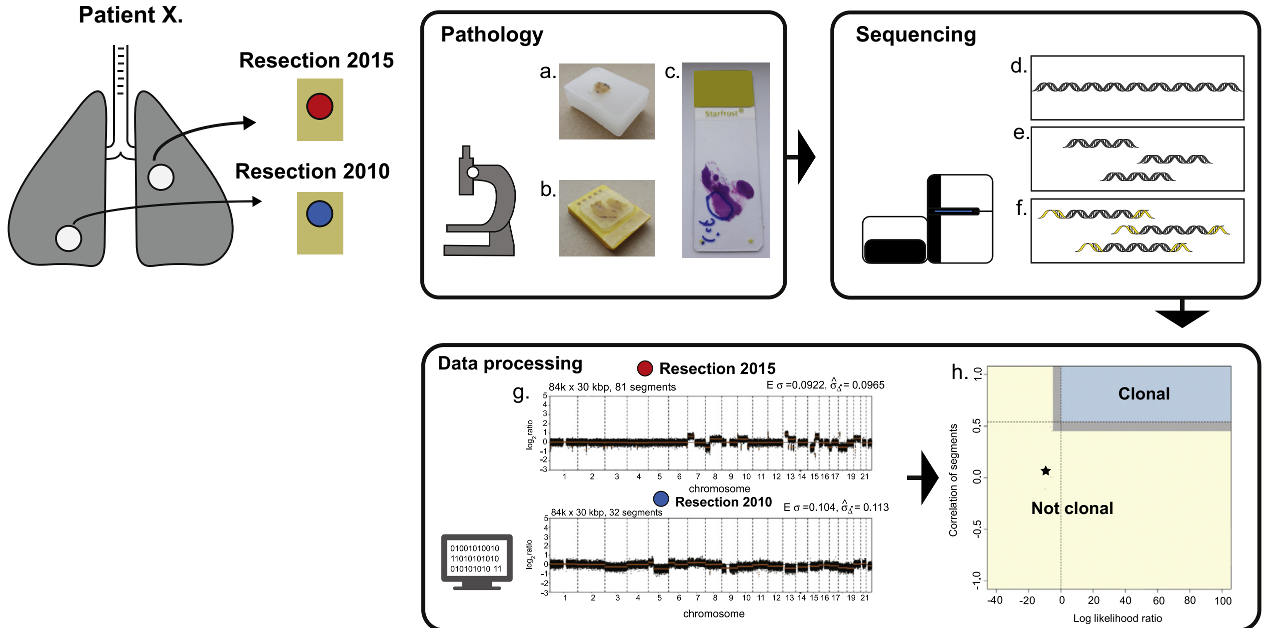


FIGURE 15.1 Workflow of clonality test. The primary resected tumor from the lung (2010) is compared with a secondary resected tumor from the lung (2015). Tissue is formalin-fixed and embedded in paraffin (A, B). Sections are cut (10 μm) and stained with hematoxylin–eosin (C) followed by demarcation of tumor tissue (D). DNA is isolated from cells scraped from the marked region (D). The isolated DNA is randomly fragmented (E) and sequencing adapters are ligated for compatibility with sequencing technology (F). The resulting sequencing data are processed by quantitative DNA sequencing to produce a genome-wide copy number profile. Log likelihoods and correlation of segments is calculated to determine the clonality score of the two samples. The clonality score is located in the lower left quadrant, which indicates the samples have no clonal relationship (H).

(Sie et al., 2014); however, there is a demand for even shorter turnaround times. Furthermore, we are dealing with FFPE specimens and the amount of tumor tissue is often limited, particularly for lung biopsies. Depending on the protocol used, the minimal required DNA to perform the assays ranges from 10 to 150 ng. High-level amplifications, such as the actionable *ERBB2* gene, can also be detected simultaneously with these panels (Hastings et al., 2015; Uzilov et al., 2016) by calculating deviations from the amplicon depth of coverage over a set of control samples (Hoogstraat et al., 2015).

15.4.3 Exome-Wide Mutation Analysis

In The Netherlands, only academic hospitals are allowed to perform hereditary genome diagnostics. To investigate the genetic cause of a congenital disease or cancer predisposition, whole-exome sequencing (WES) can be used. Currently, this is the most data-intensive and demanding assay used in routine clinical practice at the VU University Medical Center in Amsterdam. If there is only a single affected patient in the family, the preferred approach is to sequence the family trio: the mother, father, and affected child. The amount of data per casus can add up to 100 GB of data and the analysis can take days to process computationally, which makes the use of high performance clustered computing preferable. Clinical geneticists are tasked with the annotation, review, and interpretation of the reported variants, taking into account inheritance models to find the causal variant. Whereas WES is primarily designed for qualitative analysis of point mutations, it can also be used in a quantitative way, in which the copy number state of genomic regions is calculated based on the abundance of the enriched molecules of DNA. The abundance is represented by the number of reads that align to the genomic regions and needs to be corrected for variable enrichment efficiency (Magi et al., 2013). Not only high-level amplifications but also single copy number gains and losses may be detected, which are important for neuroblastomas or diffuse low-grade gliomas, for example (Hastings et al., 2015). Alternatively, an abundance of reads for genomic regions can also be calculated by taking into account "off-target reads" (Kuilman et al., 2015), which are essentially background signal and a product of the inefficiency of the enrichment protocol. The sWGS procedure described earlier is a more robust alternative. It requires a separate sequence run with a small aliquot of the intermediate sequence library obtained before the WES enrichment protocol. Because the sequence depth is limited for sWGS, additional costs are marginal (Scheinin et al., 2014).

15.4.4 Translocation Detection

The financial and practical feasibility of using deep WGS (>30× coverage) in a diagnostic setting is currently being evaluated in hereditary clinical genetics laboratories. WGS will consider up to 95% of the genome, in which WES considers up to 2%, and also reveal translocations in addition to copy number and mutations, and hence make the cytogenomic picture complete. Not only for hematologic malignancies but also for an array of solid tumors, translocations have important therapeutic consequences. For example, lung cancers that carry the echinoderm microtubule-associated protein-like 4−anaplastic lymphoma kinase translocation are directly susceptible to the inhibitor crizotinib (Kwak et al., 2010). Translocations are currently assayed in an elective fashion using immunohistochemistry, FISH, and/or PCR techniques (Hastings et al., 2015). These techniques have serious drawbacks and experience sensitivity, particularly because the exact translocation site may vary among tumors, and genes may have different translocation partners. WGS is an alternative but it is not yet within reach of the cancer cytogenomic clinic. Therefore, we and others have made efforts to target translocation NGS. The most attractive solutions would be those that include enrichment through hybrid selection panels such as WES. The BreaKmer procedure (Abo et al., 2015) makes use of WES data and would allow combination with such a mutation and copy number detection approach. Enrichment probes capture DNA molecules near a translocation site with sequence overhang into the translocation site, detected by the BreaKmer algorithm. An alternative approach to translocation detection is targeted locus amplification (de Vree et al., 2014). Although it is translocation partner independent, this technique is much less demanding with regard to the exact translocation site, which may be up to 1 Mb away from the probes designed. Yet, other than the BreaKmer, this procedure is not yet applicable to DNA isolated from FFPE.

15.5 IMPLEMENTATION OF CYTOGENOMICS IN THE CLINIC

The quality of a molecular-cytogenetic diagnostic assay is crucial to determine treatment options for patients. Predictive biomarkers are being assayed electively to reveal the susceptibility of a tumor to a particular cytotoxin or targeted therapy. Suboptimal results can lead to the selection of an inefficient therapy or may even have a detrimental effect on the well-being and survival of the patient. Prognostic biomarkers are being assayed to predict the (progression-free) survival of the patient and form an important determinant for treatment selection. For example, an 80-year-old patient can choose not to have a tumor treated that is predicted to remain indolent for several years, whereas a 20-year-old patient would be much more likely to select (taxing) treatment when faced with a similar prognosis.

When it comes to the introduction of cytogenomic assays into the clinic, the hereditary genome diagnostic laboratories in The Netherlands have experienced a number of development cycles of novel molecular techniques that are adopted and adapted from research to a clinical application. Diagnostic laboratories, however, adhere to a different set of requirements to warrant the quality, validity, reliability, and reproducibility of molecular assays. The molecular technique needs refinement and standardization to comply with diagnostic standards before it can be used in routine diagnostics (Weiss et al., 2013). Clinical molecular specialists in hereditary genome diagnostics or molecular pathology are tasked to interpret the molecular data and report the clinically relevant information back to the treating physician or tumor board. At the same time, they need to be aware of the latest developments in molecular biology technique and have an overview of the existing needs in the clinic, which are essential to continue innovation. The clinical molecular specialist will also need a level of understanding of bioinformatics and information technology (IT) infrastructure issues as data analysis and interpretation become increasingly complicated.

15.6 BIOINFORMATICS AND DATA ANALYSIS

The growing amount of data brought by NGS will make research and clinical applications increasingly reliant on a robust IT infrastructure and organization. Current academic hospital IT organizations are mainly geared toward supporting office applications and delivering robust information systems to manage the flow of patient information securely. However, often these organizations lack the ability to respond quickly to developments in the research setting and they often lack generic high-performance computing systems for researchers to process their data. In addition, the large differences in business culture between hospital and research are often a source of frustration and miscommunication. It is difficult for a single person to oversee all of these disciplines; therefore, a specialized team is required to bridge the gap in the genomics era between an ever-changing research environment, fast-paced genome diagnostics, and a generic hospital IT setting.

Most peripheral pathology laboratories in The Netherlands currently do not have sufficient IT resources to handle the large data flow generated by NGS. Such laboratories depend on turnkey solutions provided by commercial companies such as Sofia Genetics, CLC Bio, DNA star, J.S.I. Medical Systems GmbH, or DNAnexus. These companies provide software to analyze small-scale sequencing applications that focus on a

limited set of loci such as hot-spot panels or gene panels. A graphical user interface is provided to allow laboratory technicians and specialists to analyze and interpret the data. The investment required to acquire a commercial product is relatively small and alleviates the need for compute and storage infrastructure. Commercial companies provide an attractive and efficient means of adopting well-established analysis workflows in the clinic but are limited in their ability to co-develop new analysis workflows coming from research laboratories. Most of these software packages are locally installed tools and lack the ability to scale up resources when the need arises. Cartagenia, DNAnexus, and Sofia Genetics have solved this issue by providing their software as a service and use the scaling capabilities of public cloud computing providers to provide the required resources. However, this makes compliancy to security and privacy regulations a must because patient data are housed outside hospital networks.

15.7 CONCLUDING REMARKS

Although clinical genetics and molecular pathology are distinctly different disciplines, their common interest in cytogenomics calls for extensive collaborations. The laboratory of tumor genetics at the Radboud University Medical Center Nijmegen in The Netherlands has pioneered such a collaboration. They focus on germline mutations to determine the predisposition to cancer as well as on acquired somatic mutations to determine prognostic and/or predictive outcomes to treatment regimens. Their success has inspired us as well as many other academic hospitals in The Netherlands to follow in their footsteps. These efforts allow for efficient exploitation and the shared use of laboratory equipment and computing and storage infrastructure by both diagnostic and research applications for multiple departments. Moreover, co-location of NGS and bioinformatics on the one hand and research and diagnostic expertise on the other strengthen collaboration and shortens the development cycle.

A challenge that collides with the new demands of cytogenomic diagnostics is the current economic climate. In The Netherlands health care providers are encouraged to adopt a commercial and competitive model that makes them independent of government financial support. In many cases this leads to a lack of high-risk investment such as the acquisition of cutting-edge technology and IT infrastructure. This has a negative effect on the development cycle from research to clinical application. Nevertheless, cancer cytogenomics is a moving train on a track paved by hereditary cytogenomics. With the $1000 genome within reach, the penetrance of cancer NGS in daily clinical routine will only increase.

REFERENCES

Abo, R.P., et al., 2015. BreaKmer: detection of structural variation in targeted massively parallel sequencing data using kmers. Nucleic Acids Res. 43 (3), e19.

Amant, F., et al., 2015. Presymptomatic identification of cancers in pregnant women during noninvasive prenatal testing. JAMA Oncol. 1 (6), 814–819.

Best, M.G., et al., 2015. RNA-seq of tumor-educated platelets enables blood-based pan-cancer, multiclass, and molecular pathway cancer diagnostics. Cancer Cell 28 (5), 666–676.

Blow, N., 2007. Tissue preparation: tissue issues. Nature 448 (7156), 959–963.

Boleij, A., et al., 2015. RAS testing in metastatic colorectal cancer: excellent reproducibility amongst 17 Dutch pathology centers. Oncotarget 6 (17), 15681–15689.

Bollag, G., et al., 2012. Vemurafenib: the first drug approved for BRAF-mutant cancer. Nat. Rev. Drug Discov. 11 (11), 873–886.

Casparie, M., et al., 2007. Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. Cell. Oncol. 29 (1), 19–24.

Van Essen, H.F., Ylstra, B., 2012. High-resolution copy number profiling by array CGH using DNA isolated from formalin-fixed, paraffin-embedded tissues. Methods Mol. Biol. 838, 329–341.

Haan, J.C., et al., 2014. Genomic landscape of metastatic colorectal cancer. Nat. Commun. 5, 5457.

Hanahan, D., Weinberg, R.A., 2011. Hallmarks of cancer: the next generation. Cell 144 (5), 646–674.

Hastings, R.J., et al., February 2015. Guidelines for cytogenetic investigations in tumours. Eur. J. Hum. Genet. 1–8.

Hoogstraat, M., et al., 2015. Simultaneous detection of clinically relevant mutations and amplifications for routine cancer pathology. JMD 17 (1), 10–18.

Hyman, D.M., et al., 2015. Vemurafenib in multiple nonmelanoma cancers with *BRAF* V600 mutations. N. Engl. J. Med. 373.

Kuilman, T., et al., 2015. CopywriteR: DNA copy number detection from off-target sequence data. Genome Biol. 16 (1), 49.

Kuiper, J.L., et al., 2015. Transformation to a squamous cell carcinoma phenotype of an *EGFR* -mutated NSCLC patient after treatment with an EGFR-tyrosine kinase inhibitor. J. Clin. Pathol. 68 (4), 320–321.

Kwak, E.L., et al., 2010. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. N. Engl. J. Med. 363 (18), 1693–1703.

Magi, A., et al., 2013. EXCAVATOR: detecting copy number variants from whole-exome sequencing data. Genome Biol. 14 (10), R120.

Muir, P., et al., 2016. The real cost of sequencing: scaling computation to keep pace with data generation. Genome Biol. 17 (1), 53.

Ostrovnaya, I., et al., 2010. A metastasis or a second independent cancer? Evaluating the clonal origin of tumors using array copy number data. Stat. Med. 29 (15), 1608–1621.

Scheinin, I., et al., 2014. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. Genome Res. 24 (12), 1–11.

Schiffman, M., et al., 2011. Human papillomavirus testing in the prevention of cervical cancer. J. Natl. Cancer Inst. 103 (5), 368–383.

Sie, D., Ylstra, B., Heideman, D.A.M., 2014. Performance of amplicon-based next generation DNA sequencing for diagnostic gene mutation profiling in oncopathology. Cell. Oncol. 37 (5).

Smeets, S.J., et al., 2009. Genetic classification of oral and oropharyngeal carcinomas identifies subgroups with a different prognosis. Cell. Oncol. 31 (4), 291–300.

Straver, R., et al., 2014. WISECONDOR: detection of fetal aberrations from shallow sequencing maternal plasma based on a within-sample comparison scheme. Nucleic Acids Res. 42 (5), e31.

Tamminga, S., et al., 2016. Maternal plasma DNA and RNA sequencing for prenatal testing. Adv. Clin. Chem. 74.

Van Thuijl, H.F., et al., 2014. Spatial and temporal evolution of distal 10q deletion, a prognostically unfavorable event in diffuse low-grade gliomas. Genome Biol. 15.

Uzilov, A.V., et al., 2016. Development and clinical application of an integrative genomic approach to personalized cancer therapy. Genome Med. 8 (1), 62.

de Vree, P.J.P., et al., 2014. Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping. Nat. Biotechnol. 32 (10), 1019−1025.

Weiss, M.m, et al., 2013. Best practice guidelines for the use of next generation sequencing (NGS) applications in genome diagnostics: a national collaborative study of Dutch genome diagnostic laboratories. Hum. Mutat. 34 (10).

Ylstra, B., et al., 2006. BAC to the future! or oligonucleotides: a perspective for micro array comparative genomic hybridization (array CGH). Nucleic Acids Res. 34 (2), 445−450.

Chapter 16

Pharmacogenomics in Clinical Care and Drug Discovery

E. Mendrinou[1], T. Katsila[1], F. Innocenti[2], A. Squassina[3] and G.P. Patrinos[1]

[1]*University of Patras School of Health Sciences, Patras, Greece;* [2]*University of North Carolina at Chapel Hill, Chapel Hill, NC, United States;* [3]*University of Cagliari, Cagliari, Italy*

16.1 INTRODUCTION

Adverse drug effects/reactions (ADRs) account for approximately 100,000 deaths per year in the United States (Lazarou et al., 1998). It is the hope of many that the advent of precision medicine will radically minimize ADRs as well as therapeutic failures. Tremendous advances have been made in molecular biology, molecular genetics and genomics, and in the development and refinement of associated methods and technologies that exhibit a major impact on the understanding of drug action and other biologically active xenobiotics. Today, it is well established that genetic variability affects drug response and efficacy (Vesell and Page, 1968a,b). Upon the completion of the Human Genome Project (Venter et al., 2001) and alongside big data readily available online (Marsh et al., 2002), the promise of precision medicine looms large. However, its translation into clinical applications has been more challenging, taking into account the complexity of both disease (disease heterogeneity, i.e., several phenotypes in a single disease) and genetic involvement (several genes contributing to a single phenotype) as well as the difficulty in prospective validation of pharmacogenomic markers in controlled trials.

In pharmacogenomics, most research focuses on the effect of polymorphisms (changes in the genetic code that occur in at least 1% of the population) on a given drug response. A polymorphism may be a single nucleotide change in the DNA sequence, known as a single nucleotide polymorphism (SNP), an insertion or deletion of nucleotides, or a change in the number of gene copies, also known as copy number variation (see also Chapter 6). Complexity increases radically, as functional variations can be found in any area of the gene, including exons, introns, untranslated regions, upstream of the transcriptional start site, and sometimes in nearby genes (Brenner and Duggan, 2004).

This chapter aims to provide necessary information about how the discipline of genetics and genomics can influence drug discovery and development and more broadly the practice of health care. Particular emphasis will be placed on examining the role of pharmacogenomics and pharmacogenomics with regard to common complex diseases. Readers will find this chapter to be a useful source for standard definitions for some of the terms that are in wide use today, almost always sorely missing from both academic and public policy-related documents on the topic. We will also describe in detail the use of pharmacogenomics in current clinical care, the ethical and legal implications of pharmacogenomics, new omics strategies (such as pharmacometabolomics-aided pharmacogenomics and whole genome sequencing), and population pharmacogenomics. Finally, we will provide an overview of publicly available web-based genetic resources.

16.2 PHARMACOGENETICS VERSUS PHARMACOGENOMICS

Interindividual variability in drug response has ignited pharmacogenetic research (Vogel, 1959). Current drug development practices consider a relatively homogeneous group (i.e., all hypertensive subjects will respond similarly to the same medication) as the origin of patient populations, whereas specific studies are conducted when differences in drug response are anticipated, as in the case of organ (kidney and liver) impairment. Nevertheless, genetic variation may contribute an additional amount of variability to drug response (Fig. 16.1). So while drugs in Europe and the

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00016-X
Copyright © 2017 Elsevier Ltd. All rights reserved.

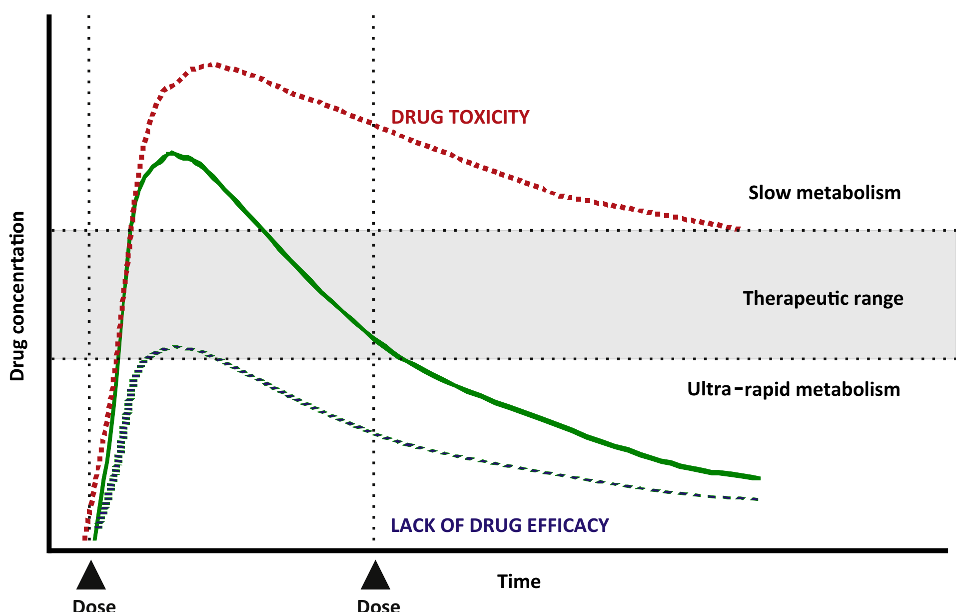


FIGURE 16.1 Graphical representation of variable drug responses. Drug metabolism normally lies within a narrow therapeutic range [depicted in green (dark gray in print versions)]. Ultra-rapid drug metabolism [depicted in blue (light gray in print versions)] results in a lack of drug efficacy, while slow drug metabolism [depicted in red (gray in print versions)] results in drug toxicity. The differences between slow and ultra-rapid metabolism have a strong genetic component, while other environmental factors also contribute to the overall differences in drug metabolism.

United States are usually tested in Caucasian patients with defined disease states, medication doses are marketed for all patients, without distinction of their geographical origin and/or considering the heterogeneity of certain diseases. Thus the contribution of genetics to the variability in drug response is what the field of pharmacogenetics and pharmacogenomics aims to answer.

The term "pharmacogenetics" was first coined by Vogel in 1959 to describe the inheritance of an aberrant drug metabolism. Pharmacogenetics has since been defined as the study of the variation in drug response as it relates to an individual's genetic makeup. Pharmacogenomics usually refers to a broader use of genome-wide association studies (GWAS) and potential complex interactions as well as alteration in gene expression that correlates to drug response. Both sciences deal with the germline heritable effects of the patient's genetic variation on drug response, and their goals are overlapping.

16.3 HISTORY OF PHARMACOGENOMICS

Prior to the discovery of DNA and the effect of genetic polymorphisms, scientists made observations on the heterogeneous nature of human drug response. Garrod (1931) defined "human biochemical individuality." His foresight into the tenants on pharmacogenomics can be seen in the following passage: "Even against chemical poisons taken by mouth, or by other channels, there are some means of defense. Every active drug is a poison, when taken in large enough doses; and in some subjects a dose which is innocuous to the majority of people has toxic effects, whereas others show exceptional tolerance to the same drug" (Garrod, 1931; Snyder et al., 1949). In 1932, the first large-scale study testing the ability of family members to taste phenylthiocarbamide was conducted and showed a clear autosomal recessive inheritance to what was then called "taste blindness" (Fox, 1932). This was also the first study to show an interethnic difference in the frequency of this phenotype. In the 1950s, various case studies supported the role of genetic variation in drug metabolism and response, such as variation in isoniazid acetylation (Hughes et al., 1954) and decreased cholinesterase activity in the action of succinylcholine (Kalow and Staron, 1957). These case studies are considered paradigms of the real possibility of pharmacologically heritable differences that may be clinically important.

In the paper by Arno Motulsky entitled "Drug reactions, enzymes and biochemical genetics" in 1957, the author outlined the genetic basis for adverse reactions to primiquine and succinylcholine as well as barbiturate-precipitated attacks and hereditary hyperbilirubinemia. Even with these landmark studies and papers, the most well-known drug-metabolizing defect discovered to affect drug metabolism was found by Matsunaga et al. (1989); the poor debrisoquine metabolizer phenotype found in 8% of Caucasian livers was attributed to a deficiency in CYP2D6,

found independently by two groups. In London, Dr. Robert Smith and some of his coworkers ingested 32 mg of debrisoquine (an obsolete antihypertensive drug); Dr. Smith experienced prolonged orthostatic hypotension hours after the dose (Smith, 1986). This effect was not seen in his coworkers. A subsequent study in medical students and three families solidified the "poor" and "extensive" metabolizer phenotype (Mahgoub et al., 1977). During this same time, Michael Eichelbaum was studying the pharmacokinetics on sparteine (an antiarrhythmic agent) when two individuals experienced nausea, diplopia, and headaches (Eichelbaum et al., 1979). It was later confirmed that the same genetic deficiency was the cause of these two independent observations. Nevertheless, the molecular genetics of this apparently simple phenotypic defect was found to be complex, with frameshift mutation, splicing defects, and a stop codon resulting in absent enzyme activity of the CYP2D6 enzyme. Today, we know that SNPs within the gene result in low enzyme activity, and gene duplication results in high enzyme activity. Because of its importance in the metabolism of several prescribed drugs and the dramatic clinical manifestation seen with aberrant metabolism, *CYP2D6* has become one of the most studied pharmacogenes.

These case studies and investigations laid the groundwork for what is now the modern pharmacogenetic discipline. The advances in the field have led to Food and Drug Administration (FDA) label changes, a guidance for industry and FDA staff on the pharmacogenomic and genetic tests for heritable markers (see http://www.fda.gov/cdrh/oivd/guidance/1549.pdf), and to unique methodology and the beginning of understanding of the genetic basis of complex diseases.

16.4 ANALYTICAL METHODS IN PHARMACOGENOMICS

Currently several approaches are employed to associate genotype (the specific nucleotide combination at a given position in the genome) with phenotype (an observable physical or biochemical characteristic), such as decreased drug clearance or increased drug toxicity as well as altered efficacy upon xenobiotic administration. Some commonly used approaches include the following: (1) candidate gene approach (i.e., phenotype/genotype or phenotype/haplotype association studies), in which SNPs are correlated to phenotype; (2) candidate pathway-based studies, which explore several genes with respect to the pharmacology of a specific drug, disease, or biological pathway for variations that may contribute to phenotypic differences; or (3) GWAS, in which thousands of SNPs throughout the genome are studied. Each of the abovementioned strategies share advantages and disadvantages. Indicatively, GWAS comprehensively survey the genetic variation in the entire genome, but their power to detect moderate associations is greatly limited, as 500,000 to 1,000,000 SNPs may be typed, resulting in multiple statistical tests and thus requiring correction for false–positive associations. In addition, GWAS require large sample sets to achieve adequate statistical power. However, these studies do provide a unique opportunity to discover novel genes not previously known to affect drug response. On the contrary, the candidate gene approach narrows the focus to a few important genes/SNPs, and therefore has higher power. Taking into account that there is a rational to the selection of these genes, the SNPs obtained are usually biologically plausible. Nevertheless, candidate gene studies may miss the real causative SNP and are limited to available knowledge.

Within the candidate-gene approach, the haplotype/phenotype approach, in which haplotypes (i.e., groups of SNPs closely linked on one chromosome and inherited as a unit) are correlated to the phenotype, holds the added advantage of assaying for an unknown variation in linkage disequilibrium (LD) with the genotyped SNPs. LD refers to the nonrandom association of alleles at two or more loci, meaning alleles that are located close to each other on the same chromosome have the tendency to occur and be inherited together. This also means that a small number of SNPs can be used to capture information about a larger region (the haplotype tagging approach). Regardless of the approach used, all positive findings need to be replicated in independent datasets to confirm their validity. There are several reviews of statistical methods in pharmacogenomics (Balding, 2006; Montana, 2006; Penrod and Moore, 2014).

16.5 PHARMACOGENOMICS IN CLINICAL SETTINGS

Although there is a plethora of scientific articles resulting from intensive genome discovery work to delineate novel and rather interesting drug response phenotypes, both in terms of drug efficacy and toxicity, there have been slightly over 120 drugs that have been approved to bear pharmacogenomics information in their labels and are routinely used in a clinical setting. The most characteristic examples of those drug/gene combinations will be outlined in the following sections.

16.5.1 *CYP2C9/VKORC1* and Warfarin

Considering the narrow therapeutic index and serious side effect profile of warfarin, its drug label has been updated to recommend genotype-guided dosing. Several studies have looked at the effect of *CYP2C9* variants on warfarin dose and on adverse events (Aithal et al., 1999; Veenstra et al., 2005a; Moridani et al., 2006), as CYP2C9 is the main enzyme of the inactivation of the

active S enantiomer of warfarin (Breckenridge et al., 1974). *CYP2C9* is part of the 4-gene *CYP2C* gene cluster that is located on chromosome 11 (Gray et al., 1995). Most studies have focused on the common functional variants in *CYP2C9*, named *CYP2C9*2* and *CYP2C9*3*. These variants require significantly lower maintenance doses of warfarin, approximately 0.85 mg for *CYP2C9*2*, 1.92 mg for *CYP2C9*3*, and 1.47 mg for both alleles, from the mean daily dose (Sanderson et al., 2005) and relate to a higher risk of serious bleeds. Various studies have shown a relationship between the genotype and the mean warfarin maintenance dosing: the dose is reduced by 18−40% in patients carrying *2C9*2/*2* alleles; by 21−49% in patients carrying *2C9*1/*3* alleles compared to *CYP2C9*1/*1* (Militaru et al., 2015). In addition, the risk of bleeding complications during warfarin treatment has been increased by 90% for *CYP2C9*2* and by 80% for *CYP2C9*3* allele variants (Wypasek et al., 2015). Both of these alleles have low allele frequency (approximately 10%) in Caucasians and even lower frequency in African Americans (2−5%) and are absent in Chinese and Japanese subjects (Wang et al., 1995; Nasu et al., 1997). The *CYP2C9*5* allele was found in healthy African American volunteers and has been hypothesized to slow the metabolism of warfarin (Dickmann et al., 2001).

In addition to *CYP2C9* genotypes, several studies have shown that variation in the *VKORC1* gene accounts for 20−30% of the observed population variation in warfarin dose (Rieder et al., 2005; Sconce et al., 2005; Veenstra et al., 2005b). Vitamin K epoxide reductase (VKOR), the therapeutic target of warfarin, is encoded by the gene *VKORC1*, in which rare mutations have been associated with warfarin resistance (Rost et al., 2004). Rieder et al. (2005) highlight the growing effort to incorporate this new gene into warfarin dosing. The investigators found 10 common SNPs in three populations that formed nine haplotypes. An association study was then conducted (excluding Asians and African Americans) to determine the effect of these haplotypes on warfarin dose. The results showed that two haplotypes were associated with low warfarin dose (haplogroup A), and three haplotypes were associated with high warfarin dose (haplogroup B). As mentioned previously, these two genes are an ideal example of the range of phenotypes seen when genetic variation affects both pharmacokinetics and the drug target. From this study we can see that patients possessing the wild type *CYP2C9* and the *VKORC1* B/B haplogroup required the highest doses, while patients possessing a variant *CYP2C9* allele and *VKORC1* haplogroup A/A required the lowest doses.

Various algorithms have been created to determine the maintenance dose of warfarin, taking into account genetic (*VKORK1* and *CYP2C9*) and nongenetic (age, weight, coadminister medications) factors (Sconce et al., 2005; Takahashi et al., 2006; Caldwell et al., 2007; Millican et al., 2007). These algorithms explain approximately 50% of the variability in warfarin dose (Sconce et al., 2005). A newly developed algorithm includes the variables of age, body weight, and *CYP2C9* and *VKORC1* genotype. This accounted for 51% of the variability in the warfarin stable dose and performed best in terms of the predicting dose (within 20% of the actual dose and the intermediate-dose group) (Cho et al., 2016). On August 16, 2007, the data from these studies led to changes in the warfarin drug label to include the consideration of *CYP2C9* and *VKORC1* genotype in warfarin dosing. No consensus was reached on how to implement such testing (i.e., which polymorphisms to include and which algorithm to recommend in the drug label), and therefore clinical implementation of this test has remained difficult.

Findings show contradictory results as to the benefit of pharmacogenomic testing to rationalize anticoagulation treatment (Pirmohamed et al., 2013). Advances in anticoagulant pharmacogenomics have the potential to improve clinical outcomes, but as with every innovation, their integration into the mainstream clinical practice is so far slow. There is definitely a need for more scientifically sound evidence from randomized clinical trials.

16.5.2 *TPMT* and 6-Mercaptopurine

6-Mercaptopurine (6-MP) and azathioprine are used in the treatment of a variety of autoimmune disorders, childhood acute lymphoblastic leukemia (ALL), rheumatoid arthritis, inflammatory bowel disease, and to prevent organ rejection. Azathioprine is converted into 6-MP, and thiopurine methyltransferase (TPMT) is involved in the methylation reactions of 6-MP, preventing the formation of its active metabolites. The clinical importance of *TPMT* pharmacogenomics has been clearly established in relation to the prediction of severe, and rarely fatal, hematological toxicity in childhood ALL patients with reduction or loss of TPMT activity. In the absence of the information on the TPMT status of patients, the drug dosage is adjusted by assessing the myelosuppressive effect of the therapy, as it is required that patients experience moderate myelosuppression. In some patients, bone marrow depression is severe, exposing them to the risk of life-threatening infections. In other patients myelosuppression is less severe, but an interruption of the therapy or dose reductions are required until the bone marrow has recovered.

Up to 28 variant alleles in the *TPMT* gene have been identified, most of them correlated with a low TPMT activity. The nonsynonymous variants *3C (719A>G), *3A (460G>A and 719A>G), and *2 (238G>C) are common (2−10%) and account for over 95% of the alleles associated with reduced TPMT activity (Wang and Weinshilboum, 2006). About 10% of ALL patients are heterozygous, and

0.3% are homozygous for the deficient alleles. As a consequence, heterozygous ALL patients are at intermediate risk of exaggerated bone marrow toxicity. In these patients, a 35% dose reduction of 6-MP has been demonstrated to be safe (Relling et al., 1999). In ALL patients, who are homozygous for defective alleles, a 90% dose reduction is required (Evans et al., 1991, 2001).

It is important to understand the impact of the heterozygous state (occurring in 10% of patients) on toxicity and survival compared to those patients lacking inactivating *TPMT* variants. Dose adjustment of 6-MP in heterozygous ALL patients preserved the efficacy of 6-MP (Relling et al., 2006), and another study has shown that *TPMT* heterozygosity has been associated with a 2.9-fold lower risk of minimal residual disease (an important prognostic factor of early response) compared to patients with no deficient *TPMT* allele (Stanulla et al., 2005). The identification of TPMT-deficient patients at the beginning of the maintenance treatment might lead to optimization of therapy by avoiding severe myelosuppression and tailoring the 6-MP dose to maintain an adequate level of exposure throughout the course of the treatment, maintaining the dose intensity of the regimen. The label of 6-MP has been revised to contain information about substantial dose reduction in *TPMT* homozygous patients (Yong and Innocenti, 2007).

16.5.3 *CYP2D6* and Tamoxifen

In the case of tamoxifen and *CYP2D6*, although research data supports a benefit of pharmacogenomic testing in dosing and/or treatment, further validation and replication are needed to solidify the case to label changes in these drugs. Tamoxifen is the standard endocrine therapy for estrogen receptor (ER)-positive breast cancer in premenopausal women. Both the parent drug and its metabolite (endoxifen and 4-hydroxy tamoxifen) have pharmacologic activity, with the metabolites of tamoxifen having 30 to 100 times greater affinity to the ER. Both of these metabolites are formed through the metabolic conversion of tamoxifen by the hepatic enzyme CYP2D6.

There are over 90 known alleles in the *CYP2D6* gene (Sims, 2008). These include polymorphisms that abolish (*4, *5), decrease (*10, *17, and *41), or increase (caused by gene duplication) function. These alleles usually classify people into the following phenotypic groups: poor metabolizers (PMs), intermediate metabolizers (IMs), extensive metabolizers (EMs), and ultrarapid metabolizers (UMs).

With respect to tamoxifen dosing, investigators have looked at the association of these alleles with endoxifen concentration. Borges et al. (2006) showed that individuals carrying at least one null (*4) of reduced function allele (*10) had similar endoxifen concentrations, while those with multiple copies of the gene had significantly higher endoxifen plasma levels. These differences in active metabolite concentrations have translated to differences in drug efficacy. Goetz et al. (2005) showed that women carrying the *CYP2D6*4/*4* genotype had shortened relapse-free time and worse disease-free survival compared to women that did not carry this allele. These findings were confirmed by Schroth et al. (2007), who showed that tamoxifen-treated women that carried a null or reduced function allele had significantly more recurrence, shorter relapse-free survival, and worse event-free survival compared to functional allele carriers. In later studies it was noted that the coadministration of CYP2D6 inhibitors (common in breast cancer therapy for the treatment of hot flashes and depression) converted the EM phenotype to a PM phenotype (Jin et al., 2005). When both of these factors were considered in treatment outcomes, patients with decreased metabolism (either by genetic variant and/or coadministered drug) had a significantly shorter time to recurrence and worse relapse-free survival (Goetz et al., 2007).

Because of these findings, the FDA advisory committee met on October 18, 2006, to determine if recommendations should be added to the drug label of tamoxifen. The consensus of the advisory committee was that the tamoxifen label should be updated to reflect the increased risk for disease recurrence in postmenopausal women with ER-positive breast cancer who have the *CYP2D6* PM genotype or women taking drugs that interact with CYP2D6. A consensus was not reached on whether *CYP2D6* testing should be recommended. The final decision from the FDA regarding the label change is pending (Goetz et al., 2008).

The current consensus on pharmacogenomic testing is that *CYP2D6* testing may be most beneficial in settings in which alternative therapies are known to be equivalent or superior to tamoxifen monotherapy, e.g., the use of aromatase inhibitors (AI) in postmenopausal women (Goetz et al., 2007). For example, in these patients, PM and IM phenotype patients should be considered for alternative therapy. In addition, the optimal sequence of hormonal therapy (tamoxifen for some duration followed by an AI) for patients considered to be EM or UM needs to be investigated (Goetz et al., 2008).

16.5.4 *UGT1A1* and Irinotecan

There are 113 known polymorphisms in the *UGT1A1* gene (Hasegawa et al., 2006), many of which have functional consequences (Tukey and Strassburg, 2000). Because of its role in bilirubin clearance, several investigators identified polymorphisms within *UGT1A1* that contribute to a common benign familial condition of decreased bilirubin glucuronidation (Gilbert's syndrome) (Bosma et al., 1995; Monaghan et al., 1996).

The *UGT1A1*28* polymorphism, which is an insertion of an extra thymine-adenine (TA) repeat in the promoter, is

associated with enzyme expression inversely associated with repeat length. This insertion polymorphism affects the TATA box upstream of the *UGT1A1* transcription initiation site, which is responsible for the binding of general transcription factor IID, which plays an important role in the initiation of transcription (Bosma et al., 1995). Therefore individuals that are homozygous for seven TA repeats, i.e., $(TA)_7$ or *UGT1A1*28*, have a 70% reduction in *UGT1A1* expression compared to those possessing the $(TA)_6$ allele (Beutler et al., 1998). *UGT1A1*28* is thought to explain up to 40% of the variability in in vitro enzyme activity of UGT1A1 (Peterkin et al., 2007). Two additional alleles at this location have been identified in persons of African ancestry: $(TA)_5$ (*UGT1A1*36*) and $(TA)_8$ (*UGT1A1*37*). Another common functional polymorphism found in exon 1 of the gene, denoted as *UGT1A1*6*, is found in Asian populations and results in the substitution of an arginine for a glycine (Aono et al., 1995). The allelic frequency of *UGT1A1*6* in Asians ranges from 18% to 23% (Akaba et al., 1998; Takeuchi et al., 2004) with a 40% reduction in enzyme activity as compared to the wild type enzyme (Jinno et al., 2003). A complete list of all known functional polymorphisms can be found at the UGT nomenclature home page (Bock et al., 2005).

Evidence points toward a more complex picture of the metabolism by UGT1A1 and functional *UGT1A* polymorphisms. Because of the structure of the *UGT1A* locus (unique exons 1 for each isoform with common exons 2 through 5), haplotypes may exist between the coding regions at this locus that affect the function of one or more of the UGT1A enzymes. A haplotype effect over the effect of *UGT1A1*28* alone at this gene locus has been postulated (Paoluzzi et al., 2004; Sai et al., 2004; Innocenti et al., 2005; Lankisch et al., 2005).

Irinotecan is approved for the treatment of metastatic colorectal cancer and is often used in combination with 5-fluorouracil, leucovorin, bevacizumab, and/or cetuximab. Irinotecan is a prodrug with therapeutic activity found in its active metabolite, 7-ethyl-10-hydoxycamptothecan (SN-38), which is a potent topoisomerase I inhibitor (Gupta et al., 1994). The SN-38 AUC has been correlated to neutropenia (Pitot et al., 2000). The major route of clearance for SN-38 is glucuronidation to SN-38 glucuronide (SN-38G) by UGT1A1, although it is possible that UGT1A9 and UGT1A7 (extrahepatic) may also have a role (Gagne et al., 2002).

The initial clinical association between SN-38 glucuronidation and irinotecan toxicity was described by Gupta et al. (1994). The evidence that this was due to variability in *UGT1A1* came from two case reports in patients with Gilbert's syndrome (Wasserman et al., 1997) as well as studies by Iyer et al. (1999) demonstrating that SN-38 is a substrate for UGT1A1. Furthermore, the extent of metabolism was inversely correlated to the number of TA repeats in the promoter region (Iyer et al., 1999). The association between *UGT1A1*28* and the risk of severe diarrhea and/or neutropenia in patients receiving irinotecan was seen in a retrospective study by Ando et al. (2000). Iyer et al. (2002) conducted the first prospective trial of irinotecan pharmacogenomics, which showed an association between *UGT1A1*28* and a decreased SN-38 glucuronidation rate. This study was amended to increase the dose of irinotecan to 350 mg/m^2 (Innocenti et al., 2004). Under the revised protocol by Innocenti et al. (2004), the investigators found a significant association of *UGT1A1*28* with an incidence of grade 4 neutropenia. Following this initial work, several studies found an association of *UGT1A1*28* and pharmacokinetic parameters, i.e., a decreased SN-38G/SN-38 area under the plasma concentration-time curve (AUC) ratio and SN-38 AUCs (Paoluzzi et al., 2004; Sai et al., 2004; Araki et al., 2006), and drug toxicity such as neutropenia (Ando et al., 2000; Rouits et al., 2004; Toffoli et al., 2006) and diarrhea (Marcuello et al., 2004; Massacesi et al., 2006). For a review on these studies, see Kim and Innocenti (2007).

This evidence led to the inclusion by the FDA of *UGT1A1*28* as a risk factor for severe neutropenia in patients receiving irinotecan (http://www.fda.gov/cder/foi/label/2005/020571s024,027,028lbl.pdf). In August 2005, the FDA approved the marketing of the Invader UGT1A1 assay (Third Wave Technologies, Madison, WI, USA) as a diagnostic test for the *UGT1A1*28* allele. *UGT1A1* genotype results can also be obtained at off-site facilities, such as Mayo Medical Laboratories or ARUP Laboratories, which process whole blood samples and return genotype data directly to the physician.

16.5.5 *HLA*B5710* and Abacavir

Pharmacogenomic effects may account for not only differential efficacy but may also contribute to the differential occurrence of adverse effects. An example of this scenario is provided by the pharmacogenomic association between *HLA*B5710* and abacavir. Abacavir is a nucleoside analogue inhibitor of reverse transcriptase used to treat HIV-1 infections. Usual formulations are abacavir alone or in combination with other antiretrovirals such as lamivudine and zidovudine. These treatments are highly effective, though 2–9% of patients receiving the drug develop hypersensitivity reactions, which result in the discontinuation of the medication (Hetherington et al., 2001). This reaction is dose-independent and usually occurs within the first 6 weeks of therapy. Symptoms include fever, rash, and gastrointestinal and respiratory symptoms. Treatment consists of the permanent cessation of abacavir and supportive care. Death may occur with a reintroduction to abacavir.

Initial indications that the hypersensitivity reaction may be genetic came during the clinical trial of abacavir. First,

only a small percentage of patients were affected, and second, a clear racial difference in risk was seen, with black patients at a lower risk for the adverse event. Using a candidate gene approach, which included genes involved in immune response and drug metabolism, *HLA*B5701* was identified as having a strong statistical association to the hypersensitivity reaction (Hetherington et al., 2002). The finding was also independently found in a Western Australian cohort (Mallal et al., 2002). A clinical study was undertaken to determine the utility of *HLA*B5701* testing. The PREDICT-1 study randomized patients to a standard of care or prospective pharmacogenomic screening prior to therapy (Mallal et al., 2008). Those patients that were screened and were shown to be *HLA*B5701* positive were excluded for abacavir regimens. The prospective screening of patients reduced the overall incidence of clinically diagnosed hypersensitivity reaction from 7.8% in the standard of care arm to 3.4% in the prescreened arm. To prevent allopurinol and abacavir hypersensitivity, *HLA-B*58:01/57:01* screening has been highly recommended prior to commencing these therapies. Therefore Nguyen et al. (2016) aimed at developing and validating a rapid and robust screening method for *HLA-B*58:01/57:01*.

The same association has not been shown in a black population, presumably because of ethnic differences in LD patterns in the major histocompatibility complex (Hughes et al., 2004). Today, pharmacogenomic information accompanies abacavir's labeling.

16.6 POPULATION DIFFERENCES IN PHARMACOGENOMICS

Interindividual differences in drug response are an important clinical issue recognized by physicians and healthcare professionals. A difference in drug response between ethnic populations was reported in the early 1920s, when differential cardiac effects of atropine were reported between Caucasians and African Americans (Paskind, 1921). Similar population-specific effects were reported with antimalarial medication used in African American soldiers (Motulsky, 1960) and pupillary dilation after various mydriatic eye drops used in Caucasians, Asians, and African Americans (Chen and Poth, 1929).

Possible causes of these differences in drug response involve extrinsic factors (diet, environmental exposure, drug–drug interactions), intrinsic factors (age, race and ethnicity, gender, organ dysfunction, genetics), or interactions between these factors (FDA Guidance for Industry Collection of Race and Ethnicity Data in Clinical Trials, September 2005b). Pharmacogenomics refers to interindividual variations in DNA sequence related to absorption, distribution, metabolism and excretion (pharmacokinetics), or drug action (pharmacodynamics), including polymorphic variations in genes that encode drug-metabolizing enzymes, drug transporters, or drug targets (receptors). Genetic variation in these genes is considered to be an important source of variability in drug response and contributes to 25–50% of inappropriate drug responses-ADRs (overtreatment) or reduced effectiveness (undertreatment) (Spear et al., 2001). Side effects are one of the main reasons for hospitalization, which in some cases can lead to death. However, ADRs can be avoided when taking into account pharmacogenomic factors that affect the pharmacokinetic and/or pharmacodynamic aspects of drug effects to the patient.

A growing number of studies describe differences in drug response as a result of genetic polymorphisms in numerous molecules that affect either the pharmacokinetic or the pharmacodynamic properties of drugs. These molecules can be considered as new pharmacogenomic biomarkers monitoring patients, either preemptively (before the start of drug administration) or during drug administration, when the therapeutic effect is lower than expected (lack of effectiveness or ADRs). It has become clear that many genetic variations in drug-metabolizing enzymes relate to interindividual differences and display differences between populations. Notably, of the numerous drug-metabolizing enzymes that exhibit pharmacogenomic diversity between individuals, at least 88% show interethnic differences in variant frequencies. Furthermore, different variations (alleles) exist in different populations.

Much of the interethnic variation we see has to do with the early human migration out of Africa into Asia and Europe (Tishkoff and Williams, 2002). The genetic variation seen outside of Africa is generally thought to be a subset of the variation found within Africa (Cavalli-Sforza and Feldman, 2003). This migration caused populations to be genetically isolated and allowed population-specific mutations to arise. Variants that cause decreased enzyme function, such as glucose-6-phosphate dehydrogenase deficiency, posed an evolutionary advantage. In this case, this variation confers protection against malaria and therefore rose in frequency in African populations (Motulsky, 1960). It is also widely recognized that populations of recent African ancestry tend to have considerably more genetic variation, i.e., more sites within the genes are likely to be polymorphic, and the extent of LD can be much lower than in populations of recent European or Asian ancestry (Tishkoff and Verrelli, 2003), meaning that the length of the haplotypes found in African Americans is significantly shorter than Caucasians and would require typing more haplotype tagging SNPs to obtain a similar assessment of the genetic variation. Therefore pharmacogenomic studies based on Caucasian subjects may not always translate into an observed affect in another population, and this should be taken into consideration when designing pharmacogenomic studies. Evidence also suggests marked differences in warfarin dosing among Caucasian populations in Europe (Mizzi et al., 2016).

The genetic definition of race or ethnicity has usually corresponded to the continental or subcontinental regions, which include the following populations: sub-Saharan Africans, Europeans, western Asians, northern Africans, eastern Asians, Polynesians and other inhabitants of Oceania, and Native Americans (Rosenberg et al., 2002; Bamshad et al., 2003). However, 5−15% of genetic variation occurs between these groups, with the remaining variation occurring within such groups (Jorde et al., 2000). This makes classification of individuals into distinct racial categories difficult and often inaccurate. This is evident in the case of people who self-identify as African Americans. This "race" has some European ancestry, as measured through genetic markers, ranging from an estimated 7% for a sample of Jamaicans to 23% for a sample of African Americans from New Orleans (Parra et al., 1998). One commentary highlights the difficulty of the categorization of Asians, as the interpretation of "Asian" might depend upon cultural definitions. Drug response and efficacy may vary widely in these populations; however, categorical distinctions in both pharmacogenomic studies and drug recommendations are rarely made (Po, 2007).

Most studies of ethnic differences in drug metabolism have focused on drugs metabolized by the cytochrome P450 family (Zhou and Liu, 2000). *CYP450* is a gene cluster involved in the metabolism of 40−60% of drugs upon daily administration, including immunosuppressants (tacrolimus, cyclosporine), chemotherapeutics, and antidepressants. The subfamily CYP3A accounts for the largest amount of CYP proteins in the human liver and metabolizes more than 50% of all drugs (Adler et al., 2009). Noteworthy, a great variation between individuals (up to 40 times) in CYP3A activity has been observed, reporting a large number of single mononucleotide polymorphisms (SNPs) in CYP3A4 and CYP3A5 isoforms. In particular, CYP3A5 is responsible for 6−99% of the total CYP3A activity worldwide (Adler et al., 2009). Therefore CYP3A5 may be an important genetic contributor to interindividual and interracial differences in CYP3A-dependent drug clearance and response (Roy et al., 2005).

The most frequent SNP with functional importance in CYP3A5 relates to an A to G transition in intron 3 at position 6986 (*3), which leads to alternative splicing, resulting in a decrease in metabolic activity. Conversely, the presence of the wild type allele (*1) is associated with high CYP3A5 catalytic activity. The frequency of the CYP3A5*3 variant differs among populations, ranging from 27−50% in the African American population, 60−70% among Asians, and up to 85−95% in individuals of Caucasian ancestry, in which the expresser variant is uncommon (Kuehl et al., 2001). This polymorphism has been associated with an increased tacrolimus dose in transplant patients (Op den Buijsch et al., 2007); as such, it may explain differences seen in organ transplantation between Caucasians and African Americans (Dirks et al., 2004). The determination of the CYP3A5 genotype would be helpful in the design of appropriate immunosuppressive treatment and lowering toxicity by predicting the doses of tacrolimus or cyclosporine required for the induction and maintenance phases in individual transplant recipients (Vannaprasaht et al., 2013). Various studies have shown that renal transplant patients who are heterozygous (CYP3A5*1/*3) require a 1.5- to 2-fold higher dose of tacrolimus to achieve the target concentration (Press et al., 2009; Dai et al., 2006; Tavira et al., 2011). An appreciation of differences in immunosuppressant pharmacokinetics and dose requirements between various ethnicities in clinical practice is expected to improve posttransplant immunosuppressive pharmacotherapy and may thus contribute to equalizing prognostic outcomes for all transplant patients (Dirks et al., 2004). It should be noted that accounting for population differences has been conducted on relatively few drugs. Therefore the gap seen in the pharmacogenomic world and that in the clinical trial and clinical practice setting is yet to be bridged.

16.7 COMPLEX PHENOTYPES

Before the completion of the Human Genome Project, progress in understanding complex disorders was slow, and insight into their causes and mechanisms remained particularly limited (Hofker et al., 2014). Most complex diseases and traits do not follow simple inheritance patterns (Mendelian). With the completion of the Human Genome Project, researchers have found that SNPs are ubiquitous, occurring approximately once in every 1000−3000 base pairs (Sachidanandam et al., 2001). However, determining which of these SNPs is responsible for variations in drug response has been a daunting task. This association process becomes increasingly complicated when more than one gene is involved in the pharmacologic effect seen. These added complexities require knowledge of not only gene−drug interactions but also the entire pathway affected by the drug. In contrast to monogenic diseases, where a single variant is sufficient to cause a disease phenotype, complex diseases represent a greater interaction between environmental and genetic factors. A complex disease phenotype cannot result from a single genetic variant. Instead, multiple variants of low penetrance at multiple loci modulate synergistically disease susceptibility. Whereas the Mendelian disorders were initially revealed using linkage analysis in affected families followed by positional cloning strategies, complex diseases were much harder to comprehend due to their non-Mendelian inheritance patterns (Hofker et al., 2014).

Common complex diseases, and thus the vast majority of what is to be clinically applied genetics, behave almost fundamentally different from rare, classic, monogenic,

Mendelian diseases: whereas in the latter the impact of the genetic variant is typically categorical in nature due to high penetrance, in the former the presence of a disease-associated genetic variant is merely of probabilistic value, raising (or lowering) the likelihood of disease occurrence to some extent but never predicting it in a black-and-white fashion. This becomes evident in a pathway-based view of the problem of drug response. While differences in drug metabolism and mutations within drug targets contribute to the variability seen, other factors, such as polymorphism in regulatory genes (i.e., transcription factors), downregulation of essential hormone mediators, and abnormalities in promoter sites, just to name a few, contribute to the drug response variability. These variables still do not take into account the effect of environmental factors. Therefore for a pharmacogenomic test to be viable in this atmosphere, all of these variables must, in some form, be accounted for.

Communicating this difference to a public that has long been misled into a perception of everything genetic being of deterministic, Mendelian quality represents a second, no less important, and difficult challenge. Unless this effort is successful (by engaging in a true dialogue with all stakeholders) in providing the basis for informed discourse and sensible decision-making on the societal level, the full potential of our deepening understanding of biology and of these technological advances will not be recognized.

Inroads into understanding the genetic basis of complex diseases have been made through the advent of the GWAS. GWAS, which aim to correlate allele frequencies of SNPs with a disease status or trait variation in a human population, have now become a standard tool in human genetics research (van der Sijde, 2014). GWAS measure and analyze DNA sequence variations from across the human genome in an effort to identify genetic risk factors for diseases that are common in the population. The ultimate goal of GWAS is to use genetic risk factors to make predictions about who is at risk and to identify the biological underpinnings of disease susceptibility for developing new prevention and treatment strategies (Bush and Moore, 2012). The strategy of GWAS is based on the common variant–common disease hypothesis (CD/CV), which states that common genetic variation must play a major role in common diseases (Hofker et al., 2014). The CD/CV hypothesis predicts that multiple, commonly occurring variants from multiple genes individually contribute a small effect on disease susceptibility but additively exert a considerable effect in the manifestation of complex diseases (Reich and Lander, 2001).

While criticized by some for being nonhypothesis driven (i.e., no a priori hypothesis is given on which gene and/or variant contributes to the disease in question), GWAS are revolutionary because they permit interrogation of the entire human genome in thousands of unrelated individuals, unconstrained by prior hypotheses regarding genetic associations. The past years have been flush with replicated findings in GWAS. These include complex diseases such as type II diabetes (Zeggini et al., 2008), obesity (Loos et al., 2008), and Crohn's disease (Rioux et al., 2007) as well as prostate (Eeles et al., 2008), breast (Easton et al., 2007), colon (Tenesa et al., 2008), and lung cancer (Hung et al., 2008). GWAS has also identified important pharmacogenomic associations, as seen in a paper that showed an association between an SNP in *SLCO1B1* (encoding an anionic-transporting polypeptide) and myopathy after statin use (a cholesterol-lowering agent) (Link et al., 2008). These investigations have led to the association of important susceptibility loci, such as the *NOD2* in Crohn's disease and *FTO* for obesity. One of the most important outcomes of these studies has been the discovery of new biological associations in genes or regions previously unrecognized to have a role in each disease. It is interesting to note that many of the associations found have a relatively small effect, with typical odd ratios of about 1.5–2.0. These studies have also found that a substantial amount of the association to disease arose from regions outside of the coding gene regions. This is not surprising given that noncoding variation may affect gene regulation and is preponderant of coding regions in the human genome. Several groups are currently working on GWAS applied to clinical pharmacogenomics, and these results are expected to be available in the public domain in the near future. To date, 1818 GWAS papers have been published, describing 12,498 associations, as listed in the GWAS Catalog of the National Human Genome Research Institute (van der Sijde et al., 2014). GWAS have reliably linked thousands of genetic variants to hundreds of complex diseases or traits (Sainani, 2015), which owned in 17 trait categories, including digestive system disease, cardiovascular disease, and immune system disease (https://www.genome.gov/gwastudies/).

Genetics have thus begun taking up the even greater challenge of the genetic dissection of complex traits, aiming at discovering genetic profiles, which will be used for the prediction of the risk of disease, prevention, or therapy, which are the priorities of personalized medicine. So, understanding the biological basis of genetic effects will play an important role in developing new pharmacologic therapies (Bush and Moore, 2012).

16.8 PHARMACOGENOMICS AND REGULATORY AGENCIES

The greatest utility of the field of clinical pharmacogenomics has been shown in oncology, as oncology drugs exhibit a narrow therapeutic index. In oncology, the two major examples of the advancing role of pharmacogenomics in the

prevention of drug toxicities are *TPMT* and 6-MP, along with *UGT1A1* and irinotecan, both of which resulted in drug label updates. In general, these changes state a recommendation, not a requirement, to test.

Approximately 10% of the drug labels approved by the FDA contain pharmacogenomic information. In the context of drug labels, this information can be classified on the basis of their specific use, such as clinical response and differentiation, risk identification, dose selection guidance, susceptibility, resistance and differential disease diagnosis, and polymorphic drug targets. In should be noted that many of the labels containing pharmacogenomic information do not provide recommendations for a specific action (i.e., genetic testing). A complete list of valid genetic biomarkers can be found on the FDA's website (http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm).

16.9 PHARMACOGENOMICS IN DRUG DEVELOPMENT

The advent of pharmacogenomics has had a mixed response in the pharmaceutical industry, with several executives expressing concern that the technology remains unproven, expensive, and disruptive to the overall revenue in drug sales (Danzon and Towse, 2002; Hopkins et al., 2006). The foreseeable benefits to the addition of pharmacogenomics into the drug development process have now become evident. Pharmaceutical companies aim to get their new drugs out to market sooner with a reduced risk of failure. A prime example of the beneficial pharmacogenomics effect on drug development processes is seen in the case of troglitazone (an oral hypoglycemic agent use to treat type II diabetes). This drug was taken off the market due to severe and sometimes fatal hepatotoxicity. A pharmacogenomic case-control study of troglitazone-induced hepatotoxicity showed a strong correlation between elevated liver enzymes and functional mutations in GSTT1 and GSTM1 (phase II metabolic enzymes; Watanabe et al., 2003). Early evaluation of these pharmacogenomic variants may have saved a therapeutically important drug. Most large pharmaceutical companies have an internal pharmacogenomics program, even though the majority serve to weed out poor candidates in the drug development process. Thus the publicly available pharmacogenomic data in the field fundamentally comes from academia. Part of the reluctance from the drug industry comes from the concern that the FDA will delay the registration process. In March 2005, steps within the FDA were taken to encourage the voluntary submission of pharmacogenomic data with guidelines emphasizing the neutral effect on the registration processes (US Department of Health and Human Services, 2005a). This has been expected to encourage the collection of pharmacogenomic data on the large phase III cohorts used during the approval process. The European Agency for the Evaluation of Medicinal Products has also created internal structures to support pharmacogenomic development with guidance to sponsors provided through briefing meetings (Agency, 2007). Yet, within Europe, German, Irish, Dutch, and United Kingdom national agencies have received few requests from sponsors to consider pharmacogenomic data.

Most pharmacogenomic investigation involves genes or pathways related to pharmacokinetics and pharmacodynamics. Interindividual differences in absorption, distribution, metabolism (with regard to both the activation of prodrugs and the inactivation of the active molecule), or excretion of the drug associate to pharmacokinetic effects. Pharmacokinetic genes are obvious targets for pharmacogenomic research given that drug levels are easily measured and, for some drugs, are correlated to drug response (i.e., therapeutic drug monitoring). Differential effects caused by variations in pharmacokinetic genes are because of the presence at the intended site of action of inappropriate concentrations of the pharmaceutical agent and/or inappropriate metabolites, resulting in toxicity and/or lack of efficacy. Pharmacogenomics, as it relates to pharmacokinetics, has been recognized as an entity for more than 100 years, going back to the observation commonly credited to Archibald Garrod that a subset of psychiatric patients treated with the hypnotic, sulphonal, developed porphyria. Since then, the underlying genetic causes for many of the previously known differences in enzymatic activity have been elucidated, most prominently with regard to the P450 enzyme family, and these have been the subject of many reviews (Evans and Relling, 2004; Ingelman-Sundberg et al., 2007; see also http://www.imm.ki.se/CYPalleles; Table 16.1). However, such effects are also reported with membrane transporters, such as in the case of the differential activity of genetic variants of the *ABCB1* gene that affect the effective intracellular concentration of antiretrovirals (Fellay et al., 2002), anticancer agents (Nooter et al., 1990), and anticonvulsants (Kwan and Brodie, 2005), or of the purine-analogue-metabolizing enzyme thiomethyl−purinetransferase (Dubinsky et al., 2000). Table 16.1 outlines several examples of phase I and phase II drug-metabolizing enzymes of pharmacogenomic relevance for drug therapy.

Pharmacodynamic effects, on the contrary, may lead to interindividual differences in drug efficacy despite the presence of appropriate concentrations of the active drug compound at the intended site of action. Genetic variation in how the target molecule or another (downstream) member of the target molecule's mechanistic pathway can respond to the medicine modulates drug efficacy. Table 16.2 summarizes examples of some of the pharmacodynamic genes that have been shown to affect drug response.

TABLE 16.1 Examples of Pharmacogenetic Effects in Pharmacokinetic Genes

Enzymes	Affected Drugs/Substances
Phase I Enzymes	
Aldehyde–dehydrogenase	Acetaldehyde
Alcohol–dehydrogenase	Ethanol
CYP1A2	Caffeine
CYP2A6	Nicotine, coumarin
CYP2C9	Phenytoin, warfarin, tolbutamide, glipizide, fluvastatin
CYP2C19	Mephenytoin, omeprazole, pantoprazole, amitriptyline, imipramine
CYP2D6	Dextromethorphan, debrisoquine, sparteine, tamoxifen, nortriptyline, clomipramine
CYP2E1	Chlorzoxazone, caffeine
CYP3A4	Erythromycin
CYP3A5	Tacrolimus, saquinavir
Serum cholinesterase	Benzoylcholine, butyrylcholine
Paraoxonase/arylesterase	Paraoxon
Phase II Enzymes	
Acetyltransferase (NAT2)	Isoniazid, sulfamethazine, caffeine, procainamide, dapsone
Dihydropyrimidine–dehydrogenase	5-Fluorouracil
Organic anion transport protein 1B1 (SLCO1B1)	Pravastatin
Thiomethyltransferase	2-Mercaptoethanol, D-penicillamine, captopril
Thiopurine–methyltransferase (TPMT)	6-Mercaptopurine, 6-thioguanine, azathioprine
UDP–glucuronosyl–transferase 1A1 (UGT1A1)	Bilirubin, irinotecan
UDP–glucuronosyl–transferase 2B7 (UGT2B7)	Oxazepam, ketoprofen, estradiol, morphine

TABLE 16.2 Examples of Pharmacogenomic Effects in Pharmacodynamic Genes

Enzyme	Affected Drugs/Disease
5-Lipoxygenase	Monolukast (asthma)
ApoE2	LDL-C level
ACE	ACE inhibitors (hypertension)
Beta 2 adrenergic receptor	Albuterol (asthma)
Beta 1 adrenergic receptor	Beta blockers (hypertension)
Corticotropin-releasing hormone	Antidepressants
Dopamine receptor	Selective serotonin reuptake inhibitors (depression)
EGFR	Gefitinib, cetuximab (cancer)
ER alpha	Bone mineral density
FCGR3A	Infliximab (Crohn's disease)
HMGCoA reductase	Statins (cholesterol-lowering drug)
NEF3	Response to antipsychotics
TNF	Anti-TNF therapy (Crohn's disease and rheumatoid arthritis)
Serotonin receptor	Selective serotonin reuptake inhibitors (depression)
VKORC1	Warfarin dose (anticoagulation)

16.10 USEFUL RESOURCES IN PHARMACOGENOMICS

The advent of pharmacogenomic studies has led to the development of several databases, by which investigators can mine valuable genotype and phenotype information. The HapMap Project still leads the way as a source of large genotype information. This project evolved naturally from the Human Genome Project, which made the sequence for the entire human genome public available. The goal of the international HapMap Project is to identify and catalog genetic similarities and differences in human beings and to develop a haplotype map of the entire human genome that can be easily utilized in genetic association studies. Four world populations/racial groups were genotyped: 30 trios (two parents and a child) from Nigeria (YRI), 30 trios from the United States and of European ancestry (CEU), 45 unrelated Japanese (JPT), and 45 unrelated Han Chinese (CHB) (see http://www.hapmap.org). These populations were chosen to broadly represent African, European, and Asian ancestry and to identify most of the common haplotypes that exist in populations worldwide. The hope is that the information in HapMap can be used to elucidate genes that affect health, disease, and individual responses to medications and environmental factors. By using this haplotype map, researchers will use the LD pattern seen in human populations in pharmacogenomic studies. However, while the HapMap is sufficient for cataloging common variation, rarer alleles are not well represented. In addition, regions with complex and large regions of LD, such as the HLA region, may not be well represented by these data.

Another publicly available resource for pharmacogenomic information is the Pharmacogenomics and Pharmacogenomics Knowledge Base (PharmGKB; http://www.pharmgkb.org). The purpose of this database is to curate information that establishes knowledge about the relationships among drugs, diseases, and genes, including their variations and gene products (Klein et al., 2001). In addition to genes annotated to the drugs they may affect, PharmGKB also contains drug-specific pathways and genes found in these pathways. While quite thorough, this database is limited in the number of genes/drugs that are included.

DruGeVar (Dalabira et al., 2014) has been developed as an online resource to triangulate drugs with genes and variants that serve as pharmacogenomic biomarkers. On the basis of such a clinical pharmacogenomics initiative, an electronic pharmacogenomics assistant would be of value. Potamias et al. developed such a system to provide personalized drug recommendations based on the genotype-to-phenotype data presented and validated by regulatory agencies. Of course, such an electronic pharmacogenomics assistant may assist biomedical research toward the identification of new gene variant biomarker entities (Potamias et al., 2014). The Frequency of Inherited Disorders database (FINDbase; http://www.findbase.org), a relational database that records frequencies of causative mutations and pharmacogenomic markers worldwide (van Baal et al., 2007), has been established to provide a simple and expandable system for worldwide population-specific mutation frequency data documentation. Since the volume of pharmacogenomic databases and resources is quite large and variable, Table 16.3 summaries a few of the widely used databases publicly available.

16.11 NEW TRENDS IN PHARMACOGENOMICS

Drug response and/or toxicity may be highly variable among individuals. Pharmacogenomics explains such clinical phenotypic differences via their association to genomic variation. It is true, though, that pharmacogenomics does not take into account environmental factors, gut microbiome, and polypharmacy. For this, a holistic approach is required. Systems biology holds the promise to delineate interindividual variability via integrated and interacting networks of genes, proteins, and biochemical reactions. A multidisciplinary, systems-based approach coupled to information technologies is vital to addressing the bulk of the biological data that is available and, eventually, to implementing precision medicine. In this context, -omics technologies play a major role.

16.11.1 Pharmacometabolomics-Aided Pharmacogenomics

Even though genomic variation is fundamental, there is increasing recognition of the limitations of pharmacogenomics, as this approach does not consider environmental influences on drug absorption, distribution, metabolism, and excretion. In an alternative, yet complementary discipline, pharmacometabolomics aims to predict and/or evaluate drug efficacy and/or toxicity on the basis of metabotypes. Metabotypes are the net result of genetic, physiological, chemical, and environmental influences and refer to a huge list of both endogenous and exogenous chemical entities, often associated with an overall understanding of metabolic dynamics per the condition of interest. Metabolomic profile alterations associated with therapy (pharmacometabolomics) are expected to be one of the hallmarks of precision medicine, as metabotypes assist toward a more accurate definition of drug response and disease heterogeneity.

In 2011, the idea of the constructive coupling of the abovementioned omics technologies arose to reinforce the identification of clinically relevant associations. Suhre et al. (2011) reported a thorough analysis of genotype-dependent metabolic phenotypes integrating a GWAS with non-targeted metabolomics. The authors identified 37 genetic loci

TABLE 16.3 Commonly Used Resources in Pharmacogenomics

Resource	Web Address	Description
dbGAP	http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gap	Developed to archive and distribute the results of studies that have investigated the interaction of genotypes and phenotypes
dbSNP	http://www.ncbi.nlm.nih.gov/projects/SNP/index.html	Central repository for both single base nucleotide substitutions and short deletion and insertion polymorphisms
ENCODE Project	http://www.genome.gov/10005107	Project to identify and characterize functional elements in the human genome
International HapMap Project	http://www.hapmap.org/index.html.en	International repository of genetic variations that can be used to find genes that affect health, disease, and individual responses to medications and environmental factors
NIEHS SNPs	http://egp.gs.washington.edu	Single nucleotide polymorphism discovery resource focused on examining the relationships between environmental exposures, interindividual sequence variations in human genes, and disease risk in US populations
Perlegen Database	http://genome.perlegen.com	Access to several human genomic data sets released by Perlegen Sciences also available in HapMap
SeattleSNPs	http://pga.mbt.washington.edu	Single nucleotide polymorphism discovery resource focused on genes involved in pathways that underlie inflammatory responses in humans
UCSC Genome Browser	http://genome.ucsc.edu	Contains reference sequences and working draft assemblies for a large collection of genomes
Wellcome Trust Case Control Consortium	http://www.wtccc.org.uk/info/access_to_data_samples.shtml	Identifies genome sequence variants influencing major causes of human morbidity and mortality through the implementation and analysis of large-scale genome-wide association studies
DruGeVar	http://drugevar.genomicmedicinealliance.org/	Only freely available database triangulating between drugs, genes, and pharmacogenomic biomarkers
FINDbase	http://www.findbase.org/	Documents the prevalence of pharmacogenomic biomarkers in several populations and ethnic groups worldwide
ePGA	http://www.epga.gr	Web service aiming to translate pharmacogenomics information into a clinically meaningful format to facilitate clinicians toward the integration of genomics into daily medical practice

associated with blood metabolite concentrations, providing new functional insights for several disease-related associations, such as cardiovascular and kidney disorders, type II diabetes, cancer, gout, venous thromboembolism, and Crohn's disease. Pharmacometabolomic data has also been efficiently merged with pharmacogenomic data via SNP imputation of metabolomic-derived pathway data, eliminating the need for initial genotyping and at the same time making broader SNP association testing possible (Abo et al., 2012). In this context, Ji et al. (2011) explored citalopram/escitalopram treatment biomarkers, following a metabolomics analysis in plasma, according to which glycine was reported to be negatively associated with treatment outcomes leading to tag SNP genotyping for genes encoding glycine synthesis, and rs10975641 (GLDC) was defined as a response biomarker in major depressive disorder patients. Katsila et al. (2016) proposed merging pharmacometabolomic and pharmacogenomic data (to address the interplay of genomic and environmental influences in autoimmune disease) with information technologies on the basis of synergy between artificial and human intelligence. Humans can detect patterns that computer algorithms may fail to recognize. On the contrary, data-intensive and cognitively complex settings and processes limit human ability. Katsila et al. (2016) provide an overview of the human pharmacometabolomics-, metabolomics-, metabonomics-, metagenomics-, and pharmacometabolomics-aided pharmacogenomics studies reported in the literature as of late 2015.

16.11.2 Whole Genome Sequencing in Pharmacogenomics

Carriers of novel, unique variants may be categorized as IMs or even PMs when the variants in question render a drug-metabolizing enzyme and/or transporter inactive or disable its expression. As such, a comprehensive view of

one's genomic variants is necessary. Even though next generation sequencing approaches [whole exome (Price et al., 2012) or whole genome sequencing (Meyerson et al., 2010)] have been adopted for pharmacogenomics testing, there are very few whole genome sequencing studies in pharmacogenomics. Ashley et al. (2010) indicated the need for a low initial dose for warfarin when they reported gene variants suggestive of clopidogrel resistance and a positive response to lipid-lowering therapy. In 2011, whole genome resequencing revealed a substantial amount of novel/uncharacterized variations (predicted to alter protein function) (Drögemöller et al., 2011). Nelson et al. (2012) sequenced 202 genes encoding drug targets to explore rare genetic variants in 14,002 individuals and concluded that human populations harbor an abundance of rare variants, many of which are deleterious and relate to disease risk. A critical analysis was performed by Drögemöller et al. (2013), focusing on the unmet needs of pharmacogenomics research on schizophrenia. The authors showed that there is a percentage of "inaccessible genome," including the CYP and HLA genes. Whole genome sequencing was exploited to identify novel and putatively causative genomic variants that affect the structure and function of 231 pharmacogenes in a large number of human genomes from various ethnic backgrounds (Mizzi et al., 2014). The advent of whole genome sequencing technology offers an outstanding potential toward the implementation of pharmacogenomics in the clinic. The development of pharmacogenomic clinical tests as well as the understanding of the genetic effects underlying interindividual variability upon drug administration depend on how sequencing technologies evolve and are integrated into the clinic. Some key actions have been taken, whereas other needs are to be met.

16.12 ETHICAL IMPLICATIONS

Several ethical considerations play a major role when the integration of pharmacogenomic information into both healthcare decisions and patient records is taken into account, especially in terms of privacy and confidentiality issues. Should genetic information be treated as any other diagnostic test, or is special protection needed for this data? Is patient consent required prior to testing for all pharmacogenomic mutations, even those used to guide dosing or therapy, or only for those that predict disease susceptibility? Many of these concerns arise from lumping pharmacogenomic tests into the category of "genetic tests." Most pharmacogenomic tests are used for routine clinical decision-making (i.e., *CYP2D6* SNP testing to determine enzyme activity to guide drug dosing), while genetic tests that are predictive of disease susceptibility may have an impact on the patient's life, family, or third party interests (i.e., *BRCA1* and *BRCA2* genotyping for the risk of early onset breast cancer). Thus not all "genetic" tests are created equal (Roses, 2000; Green and Botkin, 2003). The views on the amount of protection needed and appropriate safeguards to be used are a much-debated area (Hedgecoe, 2006). Since the intent of pharmacogenomic testing is not to predict risk or disease, some in the field feel that the need for consent and genetic counseling is unnecessary. It is true that requiring the same consent and regulatory requirements as other genetic tests may block the integration of pharmacogenomics into clinical care and may potentially deprive its benefits to patients. Others feel that patient consent and an open dialogue between patients and clinicians on the benefits and potential risks of pharmacogenomic testing would increase awareness and make patients more willing to use pharmacogenomic testing (Robertson, 2001; Schubert, 2004; van Delden et al., 2004).

16.12.1 Pharmacogenomics Legislation

Concerns such as these have been the impetus for the bill regarding discrimination and genetic testing, dubbed GINA (genetic information nondiscrimination act). This bill aims to prohibit discrimination on the basis of genetic information with respect to health insurance and employment. Such a protection encourages the public to confidently partake in genetic testing for both health and research purposes. The bill passed the US House of Representatives and the Senate and was signed into law (Organization, 2008). Similar steps have been taken in Europe with less concrete results. The 1997 European Convention on Human Rights and Biomedicine includes an article prohibiting discrimination on genetic grounds (Biomedicine, 1997). However, only four of the present European Union member states have ratified it: Spain, Portugal, Greece, and Denmark. The Universal Declaration on the Human Genome and Human Rights, adopted by United Nations Educational Scientific and Cultural Organization in 1997, though not legally binding, states that no one should be subject to discrimination on genetic grounds and that genetic information should be confidential (Biomedicine, 1997).

16.12.2 Secondary Information

A second concern regarding pharmacogenomic testing refers to secondary information. This might be the case when a patient is genotyped today for an SNP that is predictive of drug response; however, at some future date, science may be able to use this genotype to predict a predisposition to a disease (Netzer and Biller-Andorno, 2004). One example in which this may be the case is the hepatic arylamine N-acetyltransferase-2 (NAT2). This enzyme exhibits polymorphic activity in which one allele codes for fast acetylation, and several alleles code for an impaired or slow acetylation. Thus patients with a slow acetylation genotype may be at increased risk of ADRs (Tanigawara et al., 2002;

Soejima et al., 2007). N-acetylation status has also been associated with an increased risk of bladder cancer (Cartwright et al., 1982) and lung cancer (Seow et al., 1999). The same case can be made for *UGT1A1*28*. Several studies have explored the association between the number of TA repeats and breast cancer risk because of its role in estradiol glucuronidation (Guillemette et al., 2000, 2001; Sparks et al., 2004). One study in 200 African American women showed a 1.8-fold increase in breast cancer risk in premenopausal women with the $(TA)_7$ and $(TA)_8$ alleles, specifically in ER-negative breast cancers (Guillemette et al., 2000). Nevertheless, a larger case-control study in Caucasian women showed no association (Guillemette et al., 2001). Even though it becomes clear that further studies are needed, these findings may play a role in patients' willingness to be genotyped. There are no significant examples in which employment or insurance discrimination has occurred based on genetic data. Trends in public policy and legislation provide individuals with increased protection from discrimination based on genetic information obtained for clinical purposes, as reflected in the legislative action outlined in Section 16.12.1. However, because of the ever-changing nature of pharmacogenomics, further regulation may be needed as the knowledge in genetic susceptibility advances.

16.13 PUBLIC HEALTH PHARMACOGENOMICS

There have been major leaps in pharmacogenomic research, facilitated by the advent of genomic technologies. However, discoveries exhibit an asynchronous pace when the translation of research findings into the clinical setting is taken into account. To this end, the smooth incorporation of pharmacogenomics research findings into daily medical practice is hampered. This phenomenon relates more to public health genomics rather than pharmacogenomics research itself (known as Public Health Pharmacogenomics). Public Health Pharmacogenomics touches upon disciplines, such as ethics in genomics, economic evaluation in genomic medicine, genome informatics and knowledge bases, and the involvement and genetics education of the various stakeholders in the field of pharmacogenomics.

Individual genomic profiling is anticipated to provide a highly efficacious therapeutic strategy. However, a gap still exists between pharmacogenomic testing per se and the interpretation and utilization of its results in a clinical setting (also known as translation of genomic results into patient care) (Patrinos, 2010; Reydon et al., 2012). In this context, several international organizations have called for the integration of pharmacogenomics and precision medicine education into core medical curricula, focusing on the attitude and views of healthcare professionals as well as the general public. Several studies indicate that both patients and the general public desire to receive pharmacogenomic services from healthcare professionals who can confidently explain the test and interpret its implications for prescriptions. However, a gap still exists between patients' high expectations and healthcare professionals' knowledge (Mai et al., 2011, 2014). Interestingly, an awareness survey of parties (healthcare professionals, industry, academia, and government) involved in pharmacogenomics in Japan evidenced the same pattern of expectations, even though there were concerns regarding genetic privacy violations and a lack of awareness (Tamaoki et al., 2007). Proper interventions at the educational level will definitely increase understanding, facilitating the incorporation of genetics into patient care. Furthermore, the stakeholders in question appear to have a generally positive attitude, despite concerns regarding privacy issues (Mai et al., 2011, 2014).

The cost-effectiveness of pharmacogenomic tests is another crucial factor if pharmacogenomics testing is to be adopted in a clinical setting. To demonstrate its economic benefits, pharmacogenomic testing requires evidence of clinical effectiveness (Deverka et al., 2010). Reports on drug treatments for psychiatry (Perlis et al., 2009), cancer (Carlson et al., 2009), and chronic inflammatory diseases (Priest et al., 2006) have presented cost-effectiveness analyses.

No doubt, the reimbursement policies of healthcare system payers can hinder the rapid dissemination of pharmacogenomics (Ginsburg and Willard, 2009). Overall, pharmacogenomic testing may represent a resource for healthcare decision makers, leading to an increased quality of clinical care along with increased economic benefits, both for pharma and public health. Pharmacogenomic tests, however, will probably prove to be more cost-effective rather than cost-saving, or eventually cost-effective only for certain combinations of disease, treatment, and test and gene characteristics (Deverka et al., 2010; Flowers and Veenstra, 2004).

Similarly, pharmacogenomics has to deal with several issues that relate to genetic discrimination, privacy, possible implications for access to life and health insurance, and genetic stigmatization and discrimination (Robertson, 2001; Issa, 2002). In particular, when an individual is defined as a "responder" or a "nonresponder" to therapeutics, this categorization serves as the new disease label with social consequences, which involve interpersonal stigmatization or identity issues. For economic reasons, pharmaceutical companies could voluntarily ignore patients with rare or complex genetic conditions or those who are not responding to any known treatment, resulting in a deprivation of effective treatments (Rothstein and Epps, 2001).

Another crucial element of concern is the storage of genomic information in databases, considering the potential loss of confidentiality or privacy, since databases link an enormous quantity of genotypic, phenotypic, and demographic data regarding individuals (Vaszar et al., 2003).

In this regard, protection for privacy and confidentiality has to be ensured, particularly in the whole genome sequencing era, as pharmacogenomic tests may carry several types of secondary information that represent a risk of psychosocial harm or adverse insurance and/or employment implications. Moreover, particular genetic subgroups may face discrimination in accessing health care or health insurance (Smart et al., 2004).

The above are some of the key factors that relate to Public Health Pharmacogenomics, which need to be adequately addressed so that the implementation of pharmacogenomics in the clinical setting is expedited.

16.14 CONCLUSIONS AND FUTURE PERSPECTIVES

Research findings have solidified genotype associations with drug efficacy and/or toxicity in a few cases, which has been further supported by updates in drug labeling instituted by regulatory agencies. However, the clinical utility of pharmacogenetic/pharmacogenomic testing is far from optimal. Further prospective clinical studies as well as discovery research and economic evaluation data are needed to establish utility to genetic information in the clinic and drug research/development setting. The clinical implications of pharmacogenomics/pharmacogenomics also rely heavily on “genethics,” patient awareness, and the education of healthcare professionals. We stand on the cusp of precision medicine. The success of this endeavor will depend on more than knowing the genetic code. Indeed, a deeper understanding of the intricacies that regulate and underlie the code is fundamental.

ACKNOWLEDGMENTS

Part of our work has been funded by the European Commission (FP7-200754; GEN2PHEN, FP7-305444; RD-Connect; H2020-668353; UPGx), the Greek General Secretariat of Research and Technology (ΠΔΕ11_046; eMoDiA) grants to George P. Patrinos, and the Golden Helix Foundation.

REFERENCES

Abo, R., Hebbring, S., Ji, Y., Zhu, H., Zeng, Z.B., Batzler, A., Jenkins, G.D., Biernacka, J., Snyder, K., Drews, M., Fiehn, O., Fridley, B., Schaid, D., Kamatani, N., Nakamura, Y., Kubo, M., Mushiroda, T., Kaddurah-Daouk, R., Mrazek, D.A., Weinshilboum, R.M., 2012. Merging pharmacometabolomics with pharmacogenomics using ‘1000 Genomes’ single-nucleotide polymorphism imputation: selective serotonin reuptake inhibitor response pharmacogenomics. Pharmacogenet. Genomics 22, 247–253.

Adler, G., Loniewska, B., Parczewski, M., Kordek, A., Ciechanowicz, A., 2009. Frequency of common CYP3A5 gene variants in healthy Polish newborn infants. Pharmacol. Rep. 61, 947–951.

Agency, E.M., 2007. Pre-authorisation Evaluation of Medicines for Human Use. 2008.

Aithal, G.P., Day, C.P., Kesteven, P.J., Daly, A.K., 1999. Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and risk of bleeding complications. Lancet 353, 717–719.

Akaba, K., Kimura, T., Sasaki, A., Tanabe, S., Ikegami, T., Hashimoto, M., Umeda, H., Yoshida, H., Umetsu, K., Chiba, H., Yuasa, I., Hayasaka, K., 1998. Neonatal hyperbilirubinemia and mutation of the bilirubin uridine diphosphate-glucuronosyltransferase gene: a common missense mutation among Japanese, Koreans and Chinese. Biochem. Mol. Biol. Int. 46, 21–26.

Ando, Y., Saka, H., Ando, M., Sawa, T., Muro, K., Ueoka, H., Yokoyama, A., Saitoh, S., Shimokata, K., Hasegawa, Y., 2000. Polymorphisms of UDP-glucuronosyltransferase gene and irinotecan toxicity: a pharmacogenetic analysis. Cancer Res. 60, 6921–6926.

Aono, S., Adachi, Y., Uyama, E., Yamada, Y., Keino, H., Nanno, T., Koiwai, O., Sato, H., 1995. Analysis of genes for bilirubin UDP-glucuronosyltransferase in Gilbert’s syndrome. Lancet 345, 958–959.

Araki, K., Fujita, K., Ando, Y., Nagashima, F., Yamamoto, W., Endo, H., Miya, T., Kodama, K., Narabayashi, M., Sasaki, Y., 2006. Pharmacogenetic impact of polymorphisms in the coding region of the UGT1A1 gene on SN-38 glucuronidation in Japanese patients with cancer. Cancer Sci. 97, 1255–1259.

Ashley, E.A., Butte, A.J., Wheeler, M.T., 2010. Clinical assessment incorporating a personal genome. Lancet 375, 1525–1535.

Balding, D.J., 2006. A tutorial on statistical methods for population association studies. Nat. Rev. Genet. 7, 781–791.

Bamshad, M.J., Wooding, S., Watkins, W.S., Ostler, C.T., Batzer, M.A., Jorde, L.B., 2003. Human population genetic structure and inference of group membership. Am. J. Hum. Genet. 72, 578–589.

Beutler, E., Gelbart, T., Demina, A., 1998. Racial variability in the UDP-glucuronosyltransferase 1 (UGT1A1) promoter: a balanced polymorphism for regulation of bilirubin metabolism? Proc. Natl. Acad. Sci. U.S.A. 95, 8170–8174.

Biomedicine, 1997. Council of Europe: Convention on Human Rights and Biomedicine. 2008. From: http://fds.oup.com/www.oup.co.uk/pdf/bt/cassese/cases/part3/ch16/1121.pdf.

Bock, W.K., Guillemette, C., Mackenzie, C., Nebert, P.I., 2005. UGT Nomenclature Home Page. http://galien.pha.ulaval.ca/alleles/ugt1a/UGT1A1.htm.

Borges, S., Desta, Z., Li, L., Skaar, T.C., Ward, B.A., Nguyen, A., Jin, Y., Storniolo, A.M., Nikoloff, D.M., Wu, L., Hillman, G., Hayes, D.F., Stearns, V., Flockhart, D.A., 2006. Quantitative effect of CYP2D6 genotype and inhibitors on tamoxifen metabolism: implication for optimization of breast cancer treatment. Clin. Pharmacol. Ther. 80, 61–74.

Bosma, P.J., Chowdhury, J.R., Bakker, C., Gantla, S., de Boer, A., Oostra, B.A., Lindhout, D., Tytgat, G.N., Jansen, P.L., Oude Elferink, R.P., Chowdhury, N.R., 1995. The genetic basis of the reduced expression of bilirubin UDP-glucuronosyltransferase 1 in Gilbert’s syndrome. N. Engl. J. Med. 333, 1171–1175.

Breckenridge, A., Orme, M., Wesseling, H., Lewis, R.J., Gibbons, R., 1974. Pharmacokinetics and pharmacodynamics of the enantiomers of warfarin in man. Clin. Pharmacol. Ther. 15, 424–430.

Brenner, C., Duggan, D., 2004. Oncogenomics: Molecular Approaches to Cancer. Wiley-Liss, Hoboken.

Bush, W.S., Moore, J.H., 2012. Chapter 11: Genome-wide association studies. PLoS Comput. Biol. 8, e1002822.

Caldwell, M.D., Berg, R.L., Zhang, K.Q., Glurich, I., Schmelzer, J.R., Yale, S.H., Vidaillet, H.J., Burmester, J.K., 2007. Evaluation of genetic factors for warfarin dose prediction. Clin. Med. Res. 5, 8−16.

Carlson, J.J., Garrison, L.P., Ramsey, S.D., Veenstra, D.L., 2009. The potential clinical and economic outcomes of pharmacogenomic approaches to EGFR-tyrosine kinase inhibitor therapy in non-small-cell lung cancer. Value Health 12, 20−27.

Cartwright, R.A., Glashan, R.W., Rogers, H.J., Ahmad, R.A., Barham-Hall, D., Higgins, E., Kahn, M.A., 1982. Role of *N*-acetyltransferase phenotypes in bladder carcinogenesis: a pharmacogenetic epidemiological approach to bladder cancer. Lancet 2, 842−845.

Cavalli-Sforza, L.L., Feldman, M.W., 2003. The application of molecular genetic approaches to the study of human evolution. Nat. Genet. 33 (Suppl.), 266−275.

Chen, K.K., Poth, E.J., 1929. Racial differences as illustrated by the mydriatic action of cocaine, euphthalmine, and ephedrine. J. Pharmacol. Exp. Ther. 36, 363−400.

Cho, S.-M., Lee, K.-Y., Choi, J.R., Lee, K.-A., 2016. Development and comparison of warfarin dosing algorithms in stroke patients. Yonsei Med. J. 57 (3), 635−640.

Dai, Y., Hebert, M.F., Isoherranen, N., Davis, C.L., Marsh, C., Shen, D.D., Thummel, K.E., 2006. Effect of CYP3A5 polymorphism on tacrolimus metabolic clearance in vitro. Drug Metab. Dispos. 34, 836−847.

Dalabira, E., Viennas, E., Daki, E., Komianou, A., Bartsakoulia, M., Poulas, K., Katsila, T., Tzimas, G., Patrinos, G.P., 2014. DruGeVar: an online resource triangulating drugs with genes and genomic biomarkers for clinical pharmacogenomics. Public Health Genomics 17 (5−6), 265−271.

Danzon, P., Towse, A., 2002. The economics of gene therapy and of pharmacogenomics. Value Health 5, 5−13.

Deverka, P.A., Vernon, J., McLeod, H.L., 2010. Economic opportunities and challenges for pharmacogenomics. Annu. Rev. Pharmacol. Toxicol. 50, 423−437.

Dickmann, L.J., Rettie, A.E., Kneller, M.B., Kim, R.B., Wood, A.J., Stein, C.M., Wilkinson, G.R., Schwarz, U.I., 2001. Identification and functional characterization of a new CYP2C9 variant (CYP2C9*5) expressed among African Americans. Mol. Pharmacol. 60, 382−387.

Dirks, N.L., Huth, B., Yates, C.R., Meibohm, B., 2004. Pharmacokinetics of immunosuppressants: a perspective on ethnic differences. Int. J. Clin. Pharmacol. Ther. 42, 701−718.

Drögemöller, B.I., Wright, G.E.B., Niehaus, D.J.H., Emsley, R.A., Warnich, L., 2011. Whole-genome resequencing in pharmacogenomics: moving away from past disparities to globally representative applications. Pharmacogenomics 12, 1717−1728.

Drögemöller, B.I., Wright, G.E., Niehaus, D.J., Emsley, R., Warnich, L., 2013. Next-generation sequencing of pharmacogenes: a critical analysis focusing on schizophrenia treatment. Pharmacogenet. Genomics 23, 666−674.

Dubinsky, M.C., Lamothe, S., Yang, H.Y., Targan, S.R., Sinnett, D., Theoret, Y., Seidman, E.G., 2000. Pharmacogenomics and metabolite measurement for 6-mercaptopurine therapy in inflammatory bowel disease. Gastroenterology 118, 705−713.

Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D., Thompson, D., Ballinger, D.G., Struewing, J.P., Morrison, J., Field, H., Luben, R., Wareham, N., Ahmed, S., Healey, C.S., Bowman, R., Meyer, K.B., Haiman, C.A., Kolonel, L.K., Henderson, B.E., Le Marchand, L., Brennan, P., Sangrajrang, S., Gaborieau, V., Odefrey, F., Shen, C.Y., Wu, P.E., Wang, H.C., Eccles, D., Evans, D.G., Peto, J., Fletcher, O., Johnson, N., Seal, S., Stratton, M.R., Rahman, N., Chenevix-Trench, G., Bojesen, S.E., Nordestgaard, B.G., Axelsson, C.K., Garcia-Closas, M., Brinton, L., Chanock, S., Lissowska, J., Peplonska, B., Nevanlinna, H., Fagerholm, R., Eerola, H., Kang, D., Yoo, K.Y., Noh, D.Y., Ahn, S.H., Hunter, D.J., Hankinson, S.E., Cox, D.G., Hall, P., Wedren, S., Liu, J., Low, Y.L., Bogdanova, N., Schurmann, P., Dork, T., Tollenaar, R.A., Jacobi, C.E., Devilee, P., Klijn, J.G., Sigurdson, A.J., Doody, M.M., Alexander, B.H., Zhang, J., Cox, A., Brock, I.W., MacPherson, G., Reed, M.W., Couch, F.J., Goode, E.L., Olson, J.E., Meijers-Heijboer, H., van den Ouweland, A., Uitterlinden, A., Rivadeneira, F., Milne, R.L., Ribas, G., Gonzalez-Neira, A., Benitez, J., Hopper, J.L., McCredie, M., Southey, M., Giles, G.G., Schroen, C., Justenhoven, C., Brauch, H., Hamann, U., Ko, Y.D., Spurdle, A.B., Beesley, J., Chen, X., Mannermaa, A., Kosma, V.M., Kataja, V., Hartikainen, J., Day, N.E., Cox, D.R., Ponder, B.A., 2007. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447, 1087−1093.

Eeles, R.A., Kote-Jarai, Z., Giles, G.G., Olama, A.A., Guy, M., Jugurnauth, S.K., Mulholland, S., Leongamornlert, D.A., Edwards, S.M., Morrison, J., Field, H.I., Southey, M.C., Severi, G., Donovan, J.L., Hamdy, F.C., Dearnaley, D.P., Muir, K.R., Smith, C., Bagnato, M., Ardern-Jones, A.T., Hall, A.L., O'Brien, L.T., Gehr-Swain, B.N., Wilkinson, R.A., Cox, A., Lewis, S., Brown, P.M., Jhavar, S.G., Tymrakiewicz, M., Lophatananon, A., Bryant, S.L., Horwich, A., Huddart, R.A., Khoo, V.S., Parker, C.C., Woodhouse, C.J., Thompson, A., Christmas, T., Ogden, C., Fisher, C., Jamieson, C., Cooper, C.S., English, D.R., Hopper, J.L., Neal, D.E., Easton, D.F., 2008. Multiple newly identified loci associated with prostate cancer susceptibility. Nat. Genet. 40, 316−321.

Eichelbaum, M., Spannbrucker, N., Steincke, B., Dengler, H.J., 1979. Defective N-oxidation of sparteine in man: a new pharmacogenetic defect. Eur. J. Clin. Pharmacol. 16, 183−187.

Evans, W.E., Relling, M.V., 2004. Moving towards individualized medicine with pharmacogenomics. Nature 429, 464−468.

Evans, W.E., Hon, Y.Y., Bomgaars, L., Coutre, S., Holdsworth, M., Janco, R., Kalwinsky, D., Keller, F., Khatib, Z., Margolin, J., Murray, J., Quinn, J., Ravindranath, Y., Ritchey, K., Roberts, W., Rogers, Z.R., Schiff, D., Steuber, C., Tucci, F., Kornegay, N., Krynetski, E.Y., Relling, M.V., 2001. Preponderance of thiopurine *S*-methyltransferase deficiency and heterozygosity among patients intolerant to mercaptopurine or azathioprine. J. Clin. Oncol. 19, 2293−2301.

Evans, W.E., Horner, M., Chu, Y.Q., Kalwinsky, D., Roberts, W.M., 1991. Altered mercaptopurine metabolism, toxic effects, and dosage requirement in a thiopurine methyltransferase-deficient child with acute lymphocytic leukemia. J. Pediatr. 119, 985−989.

Fellay, J., Marzolini, C., Meaden, E.R., Back, D.J., Buclin, T., Chave, J.P., Decosterd, L.A., Furrer, H., Opravil, M., Pantaleo, G., Retelska, D., Ruiz, L., Schinkel, A.H., Vernazza, P., Eap, C.B., Telenti, A., 2002. Response to antiretroviral treatment in HIV-1-infected individuals with allelic variants of the multidrug resistance transporter 1: a pharmacogenomics study. Lancet 359, 30−36.

Flowers, C.R., Veenstra, D., 2004. The role of cost-effectiveness analysis in the era of pharmacogenomics. Pharmacoeconomics 22, 481−493.

Fox, A.L., 1932. The relationship between chemical constitution and taste. Proc. Natl. Acad. Sci. U. S. A. 18, 115−120.

Gagne, J.F., Montminy, V., Belanger, P., Journault, K., Gaucher, G., Guillemette, C., 2002. Common human UGT1A polymorphisms and

the altered metabolism of irinotecan active metabolite 7-ethyl-10-hydroxycamptothecin (SN-38). Mol. Pharmacol. 62, 608−617.

Garrod, A.E., 1931. The Inborn Factors in Disease: An Essay. The Clarendon Press, Oxford.

Ginsburg, G.S., Willard, H.F., 2009. Genomic and personalized medicine: foundations and applications. Transl. Res. 154, 277−287.

Goetz, M.P., Kamal, A., Ames, M.M., 2008. Tamoxifen pharmacogenomics: the role of CYP2D6 as a predictor of drug response. Clin. Pharmacol. Ther. 83, 160−166.

Goetz, M.P., Knox, S.K., Suman, V.J., Rae, J.M., Safgren, S.L., Ames, M.M., Visscher, D.W., Reynolds, C., Couch, F.J., Lingle, W.L., Weinshilboum, R.M., Fritcher, E.G., Nibbe, A.M., Desta, Z., Nguyen, A., Flockhart, D.A., Perez, E.A., Ingle, J.N., 2007. The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen. Breast Cancer Res. Treat. 101, 113−121.

Goetz, M.P., Rae, J.M., Suman, V.J., Safgren, S.L., Ames, M.M., Visscher, D.W., Reynolds, C., Couch, F.J., Lingle, W.L., Flockhart, D.A., Desta, Z., Perez, E.A., Ingle, J.N., 2005. Pharmacogenomics of tamoxifen biotransformation is associated with clinical outcomes of efficacy and hot flashes. J. Clin. Oncol. 23, 9312−9318.

Gray, I.C., Nobile, C., Muresu, R., Ford, S., Spurr, N.K., 1995. A 2.4-megabase physical map spanning the CYP2C gene cluster on chromosome 10q24. Genomics 28, 328−332.

Green, M.J., Botkin, J.R., 2003. "Genetic exceptionalism" in medicine: clarifying the differences between genetic and nongenetic tests. Ann. Intern. Med. 138, 571−575.

Guillemette, C., De Vivo, I., Hankinson, S.E., Haiman, C.A., Spiegelman, D., Housman, D.E., Hunter, D.J., 2001. Association of genetic polymorphisms in UGT1A1 with breast cancer and plasma hormone levels. Cancer Epidemiol. Biomarkers Prev. 10, 711−714.

Guillemette, C., Millikan, R.C., Newman, B., Housman, D.E., 2000. Genetic polymorphisms in uridine diphospho-glucuronosyltransferase 1A1 and association with breast cancer among African Americans. Cancer Res. 60, 950−956.

Gupta, E., Lestingi, T.M., Mick, R., Ramirez, J., Vokes, E.E., Ratain, M.J., 1994. Metabolic fate of irinotecan in humans: correlation of glucuronidation with diarrhea. Cancer Res. 54, 3723−3725.

Hasegawa, Y., Ando, Y., Ando, M., Hashimoto, N., Imaizumi, K., Shimokata, K., 2006. Pharmacogenetic approach for cancer treatment-tailored medicine in practice. Ann. N. Y. Acad. Sci. 1086, 223−232.

Hedgecoe, A.M., 2006. Context, ethics and pharmacogenomics. Stud. Hist. Philos. Biol. Biomed. Sci. 37, 566−582.

Hetherington, S., Hughes, A.R., Mosteller, M., Shortino, D., Baker, K.L., Spreen, W., Lai, E., Davies, K., Handley, A., Dow, D.J., Fling, M.E., Stocum, M., Bowman, C., Thurmond, L.M., Roses, A.D., 2002. Genetic variations in HLA-B region and hypersensitivity reactions to abacavir. Lancet 359, 1121−1122.

Hetherington, S., McGuirk, S., Powell, G., Cutrell, A., Naderer, O., Spreen, B., Lafon, S., Pearce, G., Steel, H., 2001. Hypersensitivity reactions during therapy with the nucleoside reverse transcriptase inhibitor abacavir. Clin. Ther. 23, 1603−1614.

Hofker, M.H., Fu, J., Wijmenga, C., 2014. The genome revolution and its role in understanding complex diseases. Biochim. Biophys. Acta 1842, 1889−1895.

Hopkins, M.M., Ibarreta, D., Gaisser, S., Enzing, C.M., Ryan, J., Martin, P.A., Lewis, G., Detmar, S., van den Akker-van Marle, M.E., Hedgecoe, A.M., Nightingale, P., Dreiling, M., Hartig, K.J., Vullings, W., Forde, T., 2006. Putting pharmacogenomics into practice. Nat. Biotechnol. 24, 403−410.

Hughes, A.R., Mosteller, M., Bansal, A.T., Davies, K., Haneline, S.A., Lai, E.H., Nangle, K., Scott, T., Spreen, W.R., Warren, L.L., Roses, A.D., 2004. Association of genetic variations in HLA-B region with hypersensitivity to abacavir in some, but not all, populations. Pharmacogenomics 5, 203−211.

Hughes, H.B., Biehl, J.P., Jones, A.P., Schmidt, L.H., 1954. Metabolism of isoniazid in man as related to the occurrence of peripheral neuritis. Am. Rev. Tuberc. 70, 266−273.

Hung, R.J., McKay, J.D., Gaborieau, V., Boffetta, P., Hashibe, M., Zaridze, D., Mukeria, A., Szeszenia-Dabrowska, N., Lissowska, J., Rudnai, P., Fabianova, E., Mates, D., Bencko, V., Foretova, L., Janout, V., Chen, C., Goodman, G., Field, J.K., Liloglou, T., Xinarianos, G., Cassidy, A., McLaughlin, J., Liu, G., Narod, S., Krokan, H.E., Skorpen, F., Elvestad, M.B., Hveem, K., Vatten, L., Linseisen, J., Clavel-Chapelon, F., Vineis, P., Bueno-de-Mesquita, H.B., Lund, E., Martinez, C., Bingham, S., Rasmuson, T., Hainaut, P., Riboli, E., Ahrens, W., Benhamou, S., Lagiou, P., Trichopoulos, D., Holcatova, I., Merletti, F., Kjaerheim, K., Agudo, A., Macfarlane, G., Talamini, R., Simonato, L., Lowry, R., Conway, D.I., Znaor, A., Healy, C., Zelenika, D., Boland, A., Delepine, M., Foglio, M., Lechner, D., Matsuda, F., Blanche, H., Gut, I., Heath, S., Lathrop, M., Brennan, P., 2008. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature 452, 633−637.

Ingelman-Sundberg, M., Sim, S.C., Gomez, A., Rodriguez-Antona, C., 2007. Influence of cytochrome P450 polymorphisms on drug therapies: pharmacogenetic, pharmacoepigenetic and clinical aspects. Pharmacol. Ther. 116, 496−526.

Innocenti, F., Liu, W., Chen, P., Desai, A.A., Das, S., Ratain, M.J., 2005. Haplotypes of variants in the UDP-glucuronosyltransferase1A9 and 1A1 genes. Pharmacogenet. Genomics 15, 295−301.

Innocenti, F., Undevia, S.D., Iyer, L., Chen, P.X., Das, S., Kocherginsky, M., Karrison, T., Janisch, L., Ramirez, J., Rudin, C.M., Vokes, E.E., Ratain, M.J., 2004. Genetic variants in the UDP-glucuronosyltransferase 1A1 gene predict the risk of severe neutropenia of irinotecan. J. Clin. Oncol. 22, 1382−1388.

Issa, A.M., 2002. Ethical perspectives on pharmacogenomic profiling in the drug development process. Nat. Rev. Drug Discov. 1, 300−308.

Iyer, L., Das, S., Janisch, L., Wen, M., Ramirez, J., Karrison, T., Fleming, G.F., Vokes, E.E., Schilsky, R.L., Ratain, M.J., 2002. UGT1A1*28 polymorphism as a determinant of irinotecan disposition and toxicity. Pharmacogenomics J. 2, 43−47.

Iyer, L., Hall, D., Das, S., Mortell, M.A., Ramirez, J., Kim, S., Di Rienzo, A., Ratain, M.J., 1999. Phenotype-genotype correlation of in vitro SN-38 (active metabolite of irinotecan) and bilirubin glucuronidation in human liver tissue with UGT1A1 promoter polymorphism. Clin. Pharmacol. Ther. 65, 576−582.

Ji, Y., Hebbring, S., Zhu, H., Jenkins, G.D., Biernacka, J., Snyder, K., Drews, M., Fiehn, O., Zeng, Z., Schaid, D., Mrazek, D.A., Kaddurah-Daouk, R., Weinshilboum, R.M., 2011. Glycine and a glycine dehydrogenase (GLDC) SNP as citalopram/escitalopram response biomarkers in depression: pharmacometabolomics-informed pharmacogenomics. Clin. Pharmacol. Ther. 89, 97−104.

Jin, Y., Desta, Z., Stearns, V., Ward, B., Ho, H., Lee, K.H., Skaar, T., Storniolo, A.M., Li, L., Araba, A., Blanchard, R., Nguyen, A., Ullmer, L., Hayden, J., Lemler, S., Weinshilboum, R.M., Rae, J.M.,

Hayes, D.F., Flockhart, D.A., 2005. CYP2D6 genotype, antidepressant use, and tamoxifen metabolism during adjuvant breast cancer treatment. J. Natl. Cancer Inst. 97, 30–39.

Jinno, H., Tanaka-Kagawa, T., Hanioka, N., Saeki, M., Ishida, S., Nishimura, T., Ando, M., Saito, Y., Ozawa, S., Sawada, J., 2003. Glucuronidation of 7-ethyl-10-hydroxycamptothecin (SN-38), an active metabolite of irinotecan (CPT-11), by human UGT1A1 variants, G71R, P229Q, and Y486D. Drug Metab. Dispos. 31, 108–113.

Jorde, L.B., Watkins, W.S., Bamshad, M.J., Dixon, M.E., Ricker, C.E., Seielstad, M.T., Batzer, M.A., 2000. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am. J. Hum. Genet. 66, 979–988.

Kalow, W., Staron, N., 1957. On distribution and inheritance of atypical forms of human serum cholinesterase, as indicated by dibucaine numbers. Can. J. Biochem. Physiol. 35, 1305–1320.

Katsila, T., Konstantinou, E., Lavda, I., Malakis, H., Papantoni, I., Skondra, L., Patrinos, G.P., 2016. Pharmacometabolomics-aided pharmacogenomics in autoimmune disease. EBioMedicine 5, 40–45.

Kim, T.W., Innocenti, F., 2007. Insights, challenges, and future directions in irinogenetics. Ther. Drug Monit. 29, 265–270.

Klein, T.E., Chang, J.T., Cho, M.K., Easton, K.L., Fergerson, R., Hewett, M., Lin, Z., Liu, Y., Liu, S., Oliver, D.E., Rubin, D.L., Shafa, F., Stuart, J.M., Altman, R.B., 2001. Integrating genotype and phenotype information: an overview of the PharmGKB project. Pharmacogenomics Research Network and Knowledge Base. Pharmacogenomics J. 1, 167–170.

Kuehl, P., Zhang, J., Lin, Y., Lamba, J., Assem, M., Schuetz, J., Watkins, P.B., Daly, A., Wrighton, S.A., Hall, S.D., Maurel, P., Relling, M., Brimer, C., Yasuda, K., Venkataramanan, R., Strom, S., Thummel, K., Boguski, M.S., Schuetz, E., 2001. Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression. Nat. Genet. 27, 383–391.

Kwan, P., Brodie, M.J., 2005. Potential role of drug transporters in the pathogenesis of medically intractable epilepsy. Epilepsia 46, 224–235.

Lankisch, T.O., Vogel, A., Eilermann, S., Fiebeler, A., Krone, B., Barut, A., Manns, M.P., Strassburg, C.P., 2005. Identification and characterization of a functional TATA box polymorphism of the UDP glucuronosyltransferase 1A7 gene. Mol. Pharmacol. 67, 1732–1739.

Lazarou, J., Pomeranz, B.H., Corey, P.N., 1998. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. JAMA 279, 1200–1205.

Link, E., Parish, S., Armitage, J., Bowman, L., Heath, S., Matsuda, F., Gut, I., Lathrop, M., Collins, R., 2008. SLCO1B1 variants and statin-induced myopathy–a genomewide study. N. Engl. J. Med. 359, 789–799.

Loos, R.J., Lindgren, C.M., Li, S., Wheeler, E., Zhao, J.H., Prokopenko, I., Inouye, M., Freathy, R.M., Attwood, A.P., Beckmann, J.S., Berndt, S.I., Jacobs, K.B., Chanock, S.J., Hayes, R.B., Bergmann, S., Bennett, A.J., Bingham, S.A., Bochud, M., Brown, M., Cauchi, S., Connell, J.M., Cooper, C., Smith, G.D., Day, I., Dina, C., De, S., Dermitzakis, E.T., Doney, A.S., Elliott, K.S., Elliott, P., Evans, D.M., Sadaf Farooqi, I., Froguel, P., Ghori, J., Groves, C.J., Gwilliam, R., Hadley, D., Hall, A.S., Hattersley, A.T., Hebebrand, J., Heid, I.M., Lamina, C., Gieger, C., Illig, T., Meitinger, T., Wichmann, H.E., Herrera, B., Hinney, A., Hunt, S.E., Jarvelin, M.R., Johnson, T., Jolley, J.D., Karpe, F., Keniry, A., Khaw, K.T., Luben, R.N., Mangino, M., Marchini, J., McArdle, W.L., McGinnis, R., Meyre, D., Munroe, P.B., Morris, A.D., Ness, A.R., Neville, M.J., Nica, A.C., Ong, K.K., O'Rahilly, S., Owen, K.R., Palmer, C.N., Papadakis, K., Potter, S., Pouta, A., Qi, L., Randall, J.C., Rayner, N.W., Ring, S.M., Sandhu, M.S., Scherag, A., Sims, M.A., Song, K., Soranzo, N., Speliotes, E.K., Syddall, H.E., Teichmann, S.A., Timpson, N.J., Tobias, J.H., Uda, M., Vogel, C.I., Wallace, C., Waterworth, D.M., Weedon, M.N., Willer, C.J., Wraight, Yuan, X., Zeggini, E., Hirschhorn, J.N., Strachan, D.P., Ouwehand, W.H., Caulfield, M.J., Samani, N.J., Frayling, T.M., Vollenweider, P., Waeber, G., Mooser, V., Deloukas, P., McCarthy, M.I., Wareham, N.J., Barroso, I., Jacobs, K.B., Chanock, S.J., Hayes, R.B., Lamina, C., Gieger, C., Illig, T., Meitinger, T., Wichmann, H.E., Kraft, P., Hankinson, S.E., Hunter, D.J., Hu, F.B., Lyon, H.N., Voight, B.F., Ridderstrale, M., Groop, L., Scheet, P., Sanna, S., Abecasis, G.R., Albai, G., Nagaraja, R., Schlessinger, D., Jackson, A.U., Tuomilehto, J., Collins, F.S., Boehnke, M., Mohlke, K.L., 2008. Common variants near MC4R are associated with fat mass, weight and risk of obesity. Nat. Genet. 40, 768–775.

Mahgoub, A., Idle, J.R., Dring, L.G., Lancaster, R., Smith, R.L., 1977. Polymorphic hydroxylation of Debrisoquine in man. Lancet 2, 584–586.

Mai, Y., et al., 2011. A critical view of the general public's awareness and physicians' opinion of the trends and potential pitfalls of genetic testing in Greece. Pers. Med. 8, 551–561.

Mai, Y., et al., 2014. Critical appraisal of the views of healthcare professionals with respect to pharmacogenomics and personalized medicine in Greece. Pers. Med. 11, 15–26.

Mallal, S., Nolan, D., Witt, C., Masel, G., Martin, A.M., Moore, C., Sayer, D., Castley, A., Mamotte, C., Maxwell, D., James, I., Christiansen, F.T., 2002. Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. Lancet 359, 727–732.

Mallal, S., Phillips, E., Carosi, G., Molina, J.M., Workman, C., Tomazic, J., Jagel-Guedes, E., Rugina, S., Kozyrev, O., Cid, J.F., Hay, P., Nolan, D., Hughes, S., Hughes, A., Ryan, S., Fitch, N., Thorborn, D., Benbow, A., 2008. HLA-B*5701 screening for hypersensitivity to abacavir. N. Engl. J. Med. 358, 568–579.

Marcuello, E., Altes, A., Menoyo, A., Del Rio, E., Gomez-Pardo, M., Baiget, M., 2004. UGT1A1 gene variations and irinotecan treatment in patients with metastatic colorectal cancer. Br. J. Cancer 91, 678–682.

Marsh, S., Kwok, P., McLeod, H.L., 2002. SNP databases and pharmacogenomics: great start, but a long way to go. Hum. Mutat. 20, 174–179.

Massacesi, C., Terrazzino, S., Marcucci, F., Rocchi, M.B., Lippe, P., Bisonni, R., Lombardo, M., Pilone, A., Mattioli, R., Leon, A., 2006. Uridine diphosphate glucuronosyl transferase 1A1 promoter polymorphism predicts the risk of gastrointestinal toxicity and fatigue induced by irinotecan-based chemotherapy. Cancer 106, 1007–1016.

Matsunaga, E., Zanger, U.M., Hardwick, J.P., Gelboin, H.V., Meyer, U.A., Gonzalez, F.J., 1989. The CYP2D gene subfamily: analysis of the molecular basis of the debrisoquine 4-hydroxylase deficiency in DA rats. Biochemistry 28, 7349–7355.

Meyerson, M., Gabriel, S., Getz, G., 2010. Advances in understanding cancer genomes through second-generation sequencing. Nat. Rev. Genet. 11, 685–696.

Militaru, F., Vesa, S., Pop, T., Buzoianu, A., 2015. Pharmacogenetics aspects of oral anticoagulants therapy. J. Med. Life 8 (2), 171–175.

Millican, E., Jacobsen-Lenzini, P.A., Milligan, P.E., Grosso, L., Eby, C., Deych, E., Grice, G., Clohisy, J.C., Barrack, R.L., Burnett, R.S., Voora, D., Gatchel, S., Tiemeier, A., Gage, B.F., 2007. Genetic-based dosing in orthopaedic patients beginning warfarin therapy. Blood 110, 1511−1515.

Mizzi, C., Peters, B., Mitropoulou, C., Mitropoulos, K., Katsila, T., Agarwal, M.R., et al., 2014. Personalized pharmacogenomics profiling using whole-genome sequencing. Pharmacogenomics 15, 1223−1234.

Mizzi, C., Dalabira, E., Kumuthini, J., Dzimiri, N., Balogh, I., Başak, N., Böhm, R., Borg, J., Borgiani, P., Bozina, N., Bruckmueller, H., Burzynska, B., Carracedo, A., Cascorbi, I., Deltas, C., Dolzan, V., Fenech, A., Grech, G., Kasiulevicius, V., Kádaši, L., Kučinskas, V., Khusnutdinova, E., Loukas, Y.L., Macek, Jr., M., Makukh, H., Mathijssen, R., Mitropoulos, K., Mitropoulou, C., Novelli, G., Papantoni, I., Pavlovic, S., Saglio, G., Setric, J., Stojiljkovic, M., Stubbs, A.P., Squassina, A., Torres, M., Turnovec, M., van Schaik, R.H., Voskarides, K., Wakil, S.M., Werk, A., Del Zompo, M., Zukic, B., Katsila, T., Lee, M.T., Motsinger-Rief, A., Mc Leod, H.L., van der Spek, P.J., Patrinos, G.P., 2016. A European spectrum of pharmacogenomic biomarkers: implications for clinical pharmacogenomics. PLoS One 11 (9).

Monaghan, G., Ryan, M., Seddon, R., Hume, R., Burchell, B., 1996. Genetic variation in bilirubin UPD-glucuronosyltransferase gene promoter and Gilbert's syndrome. Lancet 347, 578−581.

Montana, G., 2006. Statistical methods in genetics. Brief. Bioinform. 7, 297−308.

Moridani, M., Fu, L., Selby, R., Yun, F., Sukovic, T., Wong, B., Cole, D.E., 2006. Frequency of CYP2C9 polymorphisms affecting warfarin metabolism in a large anticoagulant clinic cohort. Clin. Biochem. 39, 606−612.

Motulsky, A.G., 1957. Drug reactions enzymes, and biochemical genetics. JAMA 165, 835−837.

Motulsky, A.G., 1960. Metabolic polymorphisms and the role of infectious diseases in human evolution. Hum. Biol. 32, 28−62.

Nasu, K., Kubota, T., Ishizaki, T., 1997. Genetic analysis of CYP2C9 polymorphism in a Japanese population. Pharmacogenomics 7, 405−409.

Nelson, M.R., Wegmann, D., Ehm, M.G., Kessner, D., St Jean, P., Verzilli, C., Shen, J., Tang, Z., Bacanu, S.A., Fraser, D., Warren, L., Aponte, J., Zawistowski, M., Liu, X., Zhang, H., Zhang, Y., Li, J., Li, Y., Li, L., Woollard, P., Topp, S., Hall, M.D., Nangle, K., Wang, J., Abecasis, G., Cardon, L.R., Zöllner, S., Whittaker, J.C., Chissoe, S.L., Novembre, J., Mooser, V., 2012. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science 337 (6090).

Netzer, C., Biller-Andorno, N., 2004. Pharmacogenetic testing, informed consent and the problem of secondary information. Bioethics 18, 344−360.

Nguyen, D.V., Vidal, C., Li, J., Fulton, R.B., Fernando, S.L., 2016. Validation of a rapid test for HLA-B*58:01/57:01 allele screening to detect individuals at risk for drug-induced hypersensitivity. Pharmacogenomics 17 (5), 473−480.

Nooter, K., Sonneveld, P., Oostrum, R., Herweijer, H., Hagenbeek, T., Valerio, D., 1990. Overexpression of the mdr1 gene in blast cells from patients with acute myelocytic leukemia is associated with decreased anthracycline accumulation that can be restored by cyclosporin-A. Int. J. Cancer 45, 263−268.

Op den Buijsch, R.A., Christiaans, M.H., Stolk, L.M., de Vries, J.E., Cheung, C.Y., Undre, N.A., van Hooff, J.P., van Dieijen-Visser, M.P., Bekers, O., 2007. Tacrolimus pharmacokinetics and pharmacogenomics: influence of adenosine triphosphate-binding cassette B1 (ABCB1) and cytochrome (CYP) 3A polymorphisms. Fundam. Clin. Pharmacol. 21, 427−435.

Organization, G.A., 2008. Genetic Information Nondiscrimination Act. 2008. From: http://www.geneticalliance.org/ws_display.asp?filter=policy.leg.nondiscrim.

Paoluzzi, L., Singh, A.S., Price, D.K., Danesi, R., Mathijssen, R.H., Verweij, J., Figg, W.D., Sparreboom, A., 2004. Influence of genetic variants in UGT1A1 and UGT1A9 on the in vivo glucuronidation of SN-38. J. Clin. Pharmacol. 44, 854−860.

Parra, E.J., Marcini, A., Akey, J., Martinson, J., Batzer, M.A., Cooper, R., Forrester, T., Allison, D.B., Deka, R., Ferrell, R.E., Shriver, M.D., 1998. Estimating African American admixture proportions by use of population-specific alleles. Am. J. Hum. Genet. 63, 1839−1851.

Paskind, H.A., 1921. Some differences in response to artopine in white and coloured races. J. Lab. Clin. Med. 7, 104−108.

Patrinos, G.P., 2010. General considerations for integrating pharmacogenomics into mainstream medical practice. Hum. Genomics 4, 371−374.

Penrod, N.M., Moore, J.H., 2014. Data science approaches to pharmacogenomics. Curr. Mol. Med. 14, 805−813.

Perlis, R.H., Patrick, A., Smoller, J.W., Wang, P.S., 2009. When is pharmacogenetic testing for antidepressant response ready for the clinic? A cost-effectiveness analysis based on data from the STAR*D study. Neuropsychopharmacology 34, 2227−2236.

Peterkin, V.C., Bauman, J.N., Goosen, T.C., Menning, L., Man, M.Z., Paulauskis, J.D., Williams, J.A., Myrand, S.P., 2007. Limited influence of UGT1A1*28 and no effect of UGT2B7*2 polymorphisms on UGT1A1 or UGT2B7 activities and protein expression in human liver microsomes. Br. J. Clin. Pharmacol. 64, 458−468.

Pirmohamed, M., et al., 2013. A randomized trial of genotype-guided dosing of warfarin. N. Engl. J. Med. 369, 2294−2303.

Pitot, H.C., Goldberg, R.M., Reid, J.M., Sloan, J.A., Skaff, P.A., Erlichman, C., Rubin, J., Burch, P.A., Adjei, A.A., Alberts, S.A., Schaaf, L.J., Elfring, G., Miller, L.L., 2000. Phase I dose-finding and pharmacokinetic trial of irinotecan hydrochloride (CPT-11) using a once-every-three-week dosing schedule for patients with advanced solid tumor malignancy. Clin. Cancer Res. 6, 2236−2244.

Po, A.L., 2007. Personalised medicine: who is an Asian? Lancet 369, 1770−1771.

Potamias, G., Lakiotaki, K., Katsila, T., Lee, M.T., Topouzis, S., Cooper, D.N., Patrinos, G.P., 2014. Deciphering next-generation pharmacogenomics: an information technology perspective. Open Biol. 4 (7) pii:140071.

Press, R.R., Ploeger, B.A., den Hartigh, J., van der Straaten, T., van Pelt, J., Danhof, M., de Fijter, J.W., Guchelaar, H.J., 2009. Explaining variability in tacrolimus pharmacokinetics to optimize early exposure in adult kidney transplant recipients. Ther. Drug Monit. 31, 187−197.

Price, M.J., Carson, A.R., Murray, S.S., Phillips, T., Janel, L., Tisch, R., et al., 2012. First pharmacogenomic analysis using whole exome sequencing to identify novel genetic determinants of clopidogrel response variability: results of the genotype information and functional testing (gift) exome study. J. Am. Coll. Cardiol. 59, E9.

Priest, V.L., et al., 2006. Pharmacoeconomic analyses of azathioprine, methotrexate and prospective pharmacogenetic testing for the

management of inflammatory bowel disease. Pharmacoeconomics 24, 767–781.

Reich, D.E., Lander, E.S., 2001. On the allelic spectrum of human disease. Trends Genet. 17, 502–510.

Relling, M.V., Hancock, M.L., Rivera, G.K., Sandlund, J.T., Ribeiro, R.C., Krynetski, E.Y., Pui, C.H., Evans, W.E., 1999. Mercaptopurine therapy intolerance and heterozygosity at the thiopurine *S*-methyltransferase gene locus. J. Natl. Cancer Inst. 91, 2001–2008.

Relling, M.V., Pui, C.H., Cheng, C., Evans, W.E., 2006. Thiopurine methyltransferase in acute lymphoblastic leukemia. Blood 107, 843–844.

Reydon, T.A., Kampourakis, K., Patrinos, G.P., 2012. Genetics, genomics and society: the responsibilities of scientists for science communication and education. Pers. Med. 9, 633–643.

Rieder, M.J., Reiner, A.P., Gage, B.F., Nickerson, D.A., Eby, C.S., McLeod, H.L., Blough, D.K., Thummel, K.E., Veenstra, D.L., Rettie, A.E., 2005. Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. N. Engl. J. Med. 352, 2285–2293.

Rioux, J.D., Xavier, R.J., Taylor, K.D., Silverberg, M.S., Goyette, P., Huett, A., Green, T., Kuballa, P., Barmada, M.M., Datta, L.W., Shugart, Y.Y., Griffiths, A.M., Targan, S.R., Ippoliti, A.F., Bernard, E.J., Mei, L., Nicolae, D.L., Regueiro, M., Schumm, L.P., Steinhart, A.H., Rotter, J.I., Duerr, R.H., Cho, J.H., Daly, M.J., Brant, S.R., 2007. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. Nat. Genet. 39, 596–604.

Robertson, J.A., 2001. Consent and privacy in pharmacogenetic testing. Nat. Genet. 28, 207–209.

Rosenberg, N.A., Pritchard, J.K., Weber, J.L., Cann, H.M., Kidd, K.K., Zhivotovsky, L.A., Feldman, M.W., 2002. Genetic structure of human populations. Science 298, 2381–2385.

Roses, A.D., 2000. Pharmacogenomics and the practice of medicine. Nature 405, 857–865.

Rost, S., Fregin, A., Ivaskevicius, V., Conzelmann, E., Hortnagel, K., Pelz, H.J., Lappegard, K., Seifried, E., Scharrer, I., Tuddenham, E.G., Muller, C.R., Strom, T.M., Oldenburg, J., 2004. Mutations in VKORC1 cause warfarin resistance and multiple coagulation factor deficiency type 2. Nature 427, 537–541.

Rothstein, M.A., Epps, P.G., 2001. Ethical and legal implications of pharmacogenomics. Nat. Rev. Genet. 2, 228–231.

Rouits, E., Boisdron-Celle, M., Dumont, A., Guerin, O., Morel, A., Gamelin, E., 2004. Relevance of different UGT1A1 polymorphisms in irinotecan-induced toxicity: a molecular and clinical study of 75 patients. Clin. Cancer Res. 10, 5151–5159.

Roy, J.N., Lajoie, J., Zijenah, L.S., Barama, A., Poirier, C., Ward, B.J., Roger, M., 2005. CYP3A5 genetic polymorphisms in different ethnic populations. Drug Metab. Dispos. 33, 884–887.

Sachidanandam, R., Weissman, D., Schmidt, S.C., Kakol, J.M., Stein, L.D., Marth, G., Sherry, S., Mullikin, J.C., Mortimore, B.J., Willey, D.L., Hunt, S.E., Cole, C.G., Coggill, P.C., Rice, C.M., Ning, Z., Rogers, J., Bentley, D.R., Kwok, P.Y., Mardis, E.R., Yeh, R.T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R.H., McPherson, J.D., Gilman, B., Schaffner, S., Van Etten, W.J., Reich, D., Higgins, J., Daly, M.J., Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M.C., Linton, L., Lander, E.S., Altshuler, D., 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409, 928–933.

Sai, K., Saeki, M., Saito, Y., Ozawa, S., Katori, N., Jinno, H., Hasegawa, R., Kaniwa, N., Sawada, J., Komamura, K., Ueno, K., Kamakura, S., Kitakaze, M., Kitamura, Y., Kamatani, N., Minami, H., Ohtsu, A., Shirao, K., Yoshida, T., Saijo, N., 2004. UGT1A1 haplotypes associated with reduced glucuronidation and increased serum bilirubin in irinotecan-administered Japanese patients with cancer. Clin. Pharmacol. Ther. 75, 501–515.

Sainani, K., January 12, 2015. Unlocking the genetics of complex diseases: GWAS and beyond. Biomed. Comput. Rev.

Sanderson, S., Emery, J., Higgins, J., 2005. CYP2C9 gene variants, drug dose, and bleeding risk in warfarin-treated patients: a HuGEnet systematic review and meta-analysis. Genet. Med. 7, 97–104.

Schroth, W., Antoniadou, L., Fritz, P., Schwab, M., Muerdter, T., Zanger, U.M., Simon, W., Eichelbaum, M., Brauch, H., 2007. Breast cancer treatment outcome with adjuvant tamoxifen relative to patient CYP2D6 and CYP2C19 genotypes. J. Clin. Oncol. 25, 5187–5193.

Schubert, L., 2004. Ethical implications of pharmacogenomics–do slippery slope arguments matter? Bioethics 18, 361–378.

Sconce, E.A., Khan, T.I., Wynne, H.A., Avery, P., Monkhouse, L., King, B.P., Wood, P., Kesteven, P., Daly, A.K., Kamali, F., 2005. The impact of CYP2C9 and VKORC1 genetic polymorphism and patient characteristics upon warfarin dose requirements: proposal for a new dosing regimen. Blood 106, 2329–2333.

Seow, A., Zhao, B., Poh, W.T., Teh, M., Eng, P., Wang, Y.T., Tan, W.C., Lee, E.J., Lee, H.P., 1999. NAT2 slow acetylator genotype is associated with increased risk of lung cancer among non-smoking Chinese women in Singapore. Carcinogenesis 20, 1877–1881.

Sims, S., May 2008. CYP2D6 Allele Nomenclature. From: http://www.cypalleles.ki.se/cyp2d6.htm.

Smart, A., Martin, P., Parker, M., 2004. Tailored medicine: whom will it fit? The ethics of patient and disease stratification. Bioethics 18, 322–342.

Smith, R.L., 1986. Introduction: human genetic variations in oxidative drug metabolism. Xenobiotica 16, 361–365.

Snyder, L.H., Clarke, H., Moore, C.V., 1949. Studies in human inheritance; further data on the linkage of the genes for sickle cells and the M-N blood types. Ohio J. Sci. 49, 32.

Soejima, M., Sugiura, T., Kawaguchi, Y., Kawamoto, M., Katsumata, Y., Takagi, K., Nakajima, A., Mitamura, T., Mimori, A., Hara, M., Kamatani, N., 2007. Association of the diplotype configuration at the *N*-acetyltransferase 2 gene with adverse events with co-trimoxazole in Japanese patients with systemic lupus erythematosus. Arthritis Res. Ther. 9, R23.

Sparks, R., Ulrich, C.M., Bigler, J., Tworoger, S.S., Yasui, Y., Rajan, K.B., Porter, P., Stanczyk, F.Z., Ballard-Barbash, R., Yuan, X., Lin, M.G., McVarish, L., Aiello, E.J., McTiernan, A., 2004. UDP-glucuronosyltransferase and sulfotransferase polymorphisms, sex hormone concentrations, and tumor receptor status in breast cancer patients. Breast Cancer Res. 6, R488–R498.

Spear, B.B., Heath-Chiozzi, M., Huff, J., 2001. Clinical application of pharmacogenomics. Trends Mol. Med. 7, 201–204.

Stanulla, M., Schaeffeler, E., Flohr, T., Cario, G., Schrauder, A., Zimmermann, M., Welte, K., Ludwig, W.D., Bartram, C.R., Zanger, U.M., Eichelbaum, M., Schrappe, M., Schwab, M., 2005. Thiopurine methyltransferase (TPMT) genotype and early treatment

response to mercaptopurine in childhood acute lymphoblastic leukemia. JAMA 293, 1485–1489.

Suhre, K., Shin, S.Y., Petersen, A.K., Mohney, R.P., Meredith, D., Wägele, B., Altmaier, E., CARDIoGRAM, Deloukas, P., Erdmann, J., Grundberg, E., Hammond, C.J., de Angelis, M.H., Kastenmüller, G., Köttgen, A., Kronenberg, F., Mangino, M., Meisinger, C., Meitinger, T., Mewes, H.W., Milburn, M.V., Prehn, C., Raffler, J., Ried, J.S., Römisch-Margl, W., Samani, N.J., Small, K.S., Wichmann, H.E., Zhai, G., Illig, T., Spector, T.D., Adamski, J., Soranzo, N., Gieger, C., 2011. Human metabolic individuality in biomedical and pharmaceutical research. Nature 477, 54–60.

Takahashi, H., Wilkinson, G.R., Nutescu, E.A., Morita, T., Ritchie, M.D., Scordo, M.G., Pengo, V., Barban, M., Padrini, R., Ieiri, I., Otsubo, K., Kashima, T., Kimura, S., Kijima, S., Echizen, H., 2006. Different contributions of polymorphisms in VKORC1 and CYP2C9 to intra- and inter-population differences in maintenance dose of warfarin in Japanese, Caucasians and African-Americans. Pharmacogenet. Genomics 16, 101–110.

Takeuchi, K., Kobayashi, Y., Tamaki, S., Ishihara, T., Maruo, Y., Araki, J., Mifuji, R., Itani, T., Kuroda, M., Sato, H., Kaito, M., Adachi, Y., 2004. Genetic polymorphisms of bilirubin uridine diphosphate-glucuronosyltransferase gene in Japanese patients with Crigler-Najjar syndrome or Gilbert's syndrome as well as in healthy Japanese subjects. J. Gastroenterol. Hepatol. 19, 1023–1028.

Tamaoki, M., Gushima, H., Tsutani, K., 2007. Awareness survey of parties involved in pharmacogenomics in Japan. Pharmacogenomics 8, 275–286.

Tanigawara, Y., Kita, T., Aoyama, N., Gobara, M., Komada, F., Sakai, T., Kasuga, M., Hatanaka, H., Sakaeda, T., Okumura, K., 2002. *N*-acetyltransferase 2 genotype-related sulfapyridine acetylation and its adverse events. Biol. Pharm. Bull. 25, 1058–1062.

Tavira, B., Coto, E., Díaz-Corte, C., Ortega, F., Arias, M., Torres, A., Díaz, J.M., Selgas, R., López-Larrea, C., Campistol, J.M., Alvarez, V., REDINREN Pharmacogenomics group, 2011. Pharmacogenomics of tacrolimus after renal transplantation: analysis of polymorphisms in genes encoding 16 drug metabolizing enzymes. Clin. Chem. Lab. Med. 49, 825–833.

Tenesa, A., Farrington, S.M., Prendergast, J.G., Porteous, M.E., Walker, M., Haq, N., Barnetson, R.A., Theodoratou, E., Cetnarskyj, R., Cartwright, N., Semple, C., Clark, A.J., Reid, F.J., Smith, L.A., Kavoussanakis, K., Koessler, T., Pharoah, P.D., Buch, S., Schafmayer, C., Tepel, J., Schreiber, S., Volzke, H., Schmidt, C.O., Hampe, J., Chang-Claude, J., Hoffmeister, M., Brenner, H., Wilkening, S., Canzian, F., Capella, G., Moreno, V., Deary, I.J., Starr, J.M., Tomlinson, I.P., Kemp, Z., Howarth, K., Carvajal-Carmona, L., Webb, E., Broderick, P., Vijayakrishnan, J., Houlston, R.S., Rennert, G., Ballinger, D., Rozek, L., Gruber, S.B., Matsuda, K., Kidokoro, T., Nakamura, Y., Zanke, B.W., Greenwood, C.M., Rangrej, J., Kustra, R., Montpetit, A., Hudson, T.J., Gallinger, S., Campbell, H., Dunlop, M.G., 2008. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat. Genet. 40, 631–637.

Tishkoff, S.A., Verrelli, B.C., 2003. Patterns of human genetic diversity: implications for human evolutionary history and disease. Annu. Rev. Genomics Hum. Genet. 4, 293–340.

Tishkoff, S.A., Williams, S.M., 2002. Genetic analysis of African populations: human evolution and complex disease. Nat. Rev. Genet. 3, 611–621.

Toffoli, G., Cecchin, E., Corona, G., Russo, A., Buonadonna, A., D'Andrea, M., Pasetto, L.M., Pessa, S., Errante, D., De Pangher, V., Giusto, M., Medici, M., Gaion, F., Sandri, P., Galligioni, E., Bonura, S., Boccalon, M., Biason, P., Frustaci, S., 2006. The role of UGT1A1*28 polymorphism in the pharmacodynamics and pharmacokinetics of irinotecan in patients with metastatic colorectal cancer. J. Clin. Oncol. 24, 3061–3068.

Tukey, R.H., Strassburg, C.P., 2000. Human UDP-glucuronosyltransferases: metabolism, expression, and disease. Annu. Rev. Pharmacol. Toxicol. 40, 581–616.

US Department of Health and Human Services, F.D.A., 2005a. Guidance for Industry: Pharmacogenomic 2008.

US Department of Health and Human Services, F.D.A., 2005b. Guidance for Industry: Collection of Race and Ethnicity Data in Clinical Trials.

Vannaprasaht, S., Reungjui, S., Supanya, D., Sirivongs, D., Pongskul, C., Avihingsanon, Y., Tassaneeyakul, W., 2013. Personalized tacrolimus doses determined by CYP3A5 genotype for induction and maintenance phases of kidney transplantation. Clin. Ther. 35, 1762–1769.

van Baal, S., Kaimakis, P., Phommarinh, M., Koumbi, D., Cuppens, H., Riccardino, F., Macek Jr., M., Scriver, C.R., Patrinos, G.P., 2007. *FINDbase*: a relational database recording frequencies of genetic defects leading to inherited disorders worldwide. Nucleic Acids Res. 35, D690–D695.

van Delden, J., Bolt, I., Kalis, A., Derijks, J., Leufkens, H., 2004. Tailor-made pharmacotherapy: future developments and ethical challenges in the field of pharmacogenomics. Bioethics 18, 303–321.

van der Sijde, M.R., Ng, A., Fu, J., 2014. Systems genetics: from GWAS to disease pathways. Biochim. Biophys. Acta 1842, 1903–1909.

Vaszar, L.T., Cho, M.K., Raffin, T.A., 2003. Privacy issues in personalized medicine. Pharmacogenomics 4, 107–112.

Veenstra, D.L., Blough, D.K., Higashi, M.K., Farin, F.M., Srinouanprachan, S., Rieder, M.J., Rettie, A.E., 2005a. CYP2C9 haplotype structure in European American warfarin patients and association with clinical outcomes. Clin. Pharmacol. Ther. 77, 353–364.

Veenstra, D.L., You, J.H., Rieder, M.J., Farin, F.M., Wilkerson, H.W., Blough, D.K., Cheng, G., Rettie, A.E., 2005b. Association of Vitamin K epoxide reductase complex 1 (VKORC1) variants with warfarin dose in a Hong Kong Chinese patient population. Pharmacogenet. Genomics 15, 687–691.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S.,

Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigo, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., Zhu, X., 2001. The sequence of the human genome. Science 291, 1304−1351.

Vesell, E.S., Page, J.G., 1968a. Genetic control of drug levels in man: antipyrine. Science 161, 72−73.

Vesell, E.S., Page, J.G., 1968b. Genetic control of drug levels in man: phenylbutazone. Science 159, 1479−1480.

Vogel, F., 1959. Moderne Probleme der Humangenetic. Ergeb. Inn. Med. Kinderheilkd. 12, 65−126.

Wang, L., Weinshilboum, R., 2006. Thiopurine *S*-methyltransferase pharmacogenomics: insights, challenges and future directions. Oncogene 25, 1629−1638.

Wang, S.L., Huang, J., Lai, M.D., Tsai, J.J., 1995. Detection of CYP2C9 polymorphism based on the polymerase chain reaction in Chinese. Pharmacogenomics 5, 37−42.

Wasserman, E., Myara, A., Lokiec, F., Goldwasser, F., Trivin, F., Mahjoubi, M., Misset, J.L., Cvitkovic, E., 1997. Severe CPT-11 toxicity in patients with Gilbert's syndrome: two case reports. Ann. Oncol. 8, 1049−1051.

Watanabe, I., Tomita, A., Shimizu, M., Sugawara, M., Yasumo, H., Koishi, R., Takahashi, T., Miyoshi, K., Nakamura, K., Izumi, T., Matsushita, Y., Furukawa, H., Haruyama, H., Koga, T., 2003. A study to survey susceptible genetic factors responsible for troglitazone-associated hepatotoxicity in Japanese patients with type 2 diabetes mellitus. Clin. Pharmacol. Ther. 73, 435−455.

Wypasek, E., Cieśla, M., Suder, B., Janik, Ł., Sadowski, J., Undas, A., 2015. CYP2C9 polymorphism and unstable anticoagulation with warfarin in patients within the first 3 months following heart valve replacement. Adv. Clin. Exp. Med. 24 (4), 607−614.

Yong, W.P., Innocenti, F., 2007. Translation of pharmacogenetic knowledge into cancer therapeutics. Clin. Adv. Hematol. Oncol. 5, 698−706.

Zeggini, E., Scott, L.J., Saxena, R., Voight, B.F., Marchini, J.L., Hu, T., de Bakker, P.I., Abecasis, G.R., Almgren, P., Andersen, G., Ardlie, K., Bostrom, K.B., Bergman, R.N., Bonnycastle, L.L., Borch-Johnsen, K., Burtt, N.P., Chen, H., Chines, P.S., Daly, M.J., Deodhar, P., Ding, C.J., Doney, A.S., Duren, W.L., Elliott, K.S., Erdos, M.R., Frayling, T.M., Freathy, R.M., Gianniny, L., Grallert, H., Grarup, N., Groves, C.J., Guiducci, C., Hansen, T., Herder, C., Hitman, G.A., Hughes, T.E., Isomaa, B., Jackson, A.U., Jorgensen, T., Kong, A., Kubalanza, K., Kuruvilla, F.G., Kuusisto, J., Langenberg, C., Lango, H., Lauritzen, T., Li, Y., Lindgren, C.M., Lyssenko, V., Marvelle, A.F., Meisinger, C., Midthjell, K., Mohlke, K.L., Morken, M.A., Morris, A.D., Narisu, N., Nilsson, P., Owen, K.R., Palmer, C.N., Payne, F., Perry, J.R., Pettersen, E., Platou, C., Prokopenko, I., Qi, L., Qin, L., Rayner, N.W., Rees, M., Roix, J.J., Sandbaek, A., Shields, B., Sjogren, M., Steinthorsdottir, V., Stringham, H.M., Swift, A.J., Thorleifsson, G., Thorsteinsdottir, U., Timpson, N.J., Tuomi, T., Tuomilehto, J., Walker, M., Watanabe, R.M., Weedon, M.N., Willer, C.J., Illig, T., Hveem, K., Hu, F.B., Laakso, M., Stefansson, K., Pedersen, O., Wareham, N.J., Barroso, I., Hattersley, A.T., Collins, F.S., Groop, L., McCarthy, M.I., Boehnke, M., Altshuler, D., 2008. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat. Genet. 40, 638−645.

Zhou, H.H., Liu, Z.Q., 2000. Ethnic differences in drug metabolism. Clin. Chem. Lab. Med. 38, 899−903.

Chapter 17

Nutrigenomics: Integrating Genomic Approaches Into Nutrition Research

M.P.G. Barnett[1,2] and L.R. Ferguson[2,3]

[1]AgResearch Limited, Palmerston North, New Zealand; [2]Nutrigenomics New Zealand, Auckland, New Zealand; [3]The University of Auckland, Auckland, New Zealand

17.1 INTRODUCTION

While a number of formal definitions exist, in essence nutrigenomics (sometimes called nutritional genomics) considers the interactions between foods or dietary supplements and an individual's genome and the consequent downstream effects on his or her phenotype. The field has the potential to provide tailored nutritional advice or develop specialist food products for populations or for individuals. It recognizes that appropriate dietary advice for one individual may be inappropriate or actually harmful to another. The potential is comparable to that of its sister field of pharmacogenomics, which considers individualized drug therapy. The Nutrigenomics New Zealand model (http://www.nutrigenomics.org.nz) provides an approach to integrating the various available technologies for systems biology, ultimately directed toward the development of novel and targeted foods (Ferguson, 2013; Marlow et al., 2013; Barnett et al., 2014). However, this example also provides a cohesive and integrated model that can be applied to other research objectives. Inflammatory bowel disease (IBD), specifically Crohn's disease (CD), has been used as proof of principle.

One of the many challenges in the field is combining the new developments in systems biology modeling with classical nutritional approaches to develop a more comprehensive model, which takes all relevant mechanisms into account, including phenotype (Evelo et al., 2011; Fan et al., 2013; Polytarchou et al., 2014). Nutrigenomics offers a multilevel systems biology model that is both flexible and rigorous. Because some of the newer tools provide a sensitivity that has been previously lacking, this field can be applied to the maintenance of good health as well as to the amelioration of disease states.

A necessary initial step establishes the nature of genetic variations relating to the disease or health endpoint of interest, using the techniques of genetic epidemiology (Austin et al., 2012). While many studies have measured variation primarily in the form of single nucleotide polymorphisms (SNPs) and haplotypes, an increasing number of studies are considering copy number variants (CNVs) (Abel and Duncavage, 2013). Accurate dietary assessment is also essential (Boeing, 2013; Fallaize et al., 2014; Forster et al., 2014). The information generated by these two approaches can be used as a basis for the design of a cell line or other high-throughput model that mimics the genotype and allows large numbers of food fractions or known dietary components to be tested for their ability to overcome the phenotypic effects of the variant SNP or copy number (Nasef et al., 2014; Martin et al., 2015). Food components thus identified are subjected to further testing in both in vitro and in vivo models. Animal models permit the study of defined nutrients, foods, and food components on the expression of disease, health, or performance-related genes. Food component-induced changes in gene and protein expression, metabolite levels, or intestinal microbial populations are measured in appropriate tissues collected from 2xn mouse experiments using transcriptomics (Barnett et al., 2010), proteomics (Cooney et al., 2012; Barnett et al., 2013), metabolomics (Lin et al., 2010; Lin et al., 2011), and microbiome analysis (Bassett et al., 2015) techniques. These changes can then be used to build an understanding of how foods impact complex biological systems to influence disease, health, or performance. As well as testing whether any protective properties of foods identified in vitro can be extrapolated to the more complex in vivo situations, these studies enable the identification of biomarkers, for example, protein biomarkers

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00017-1
Copyright © 2017 Elsevier Ltd. All rights reserved.

for the early and accurate diagnosis of IBD (Bennike et al., 2014; Park et al., 2014). These are then used as surrogate endpoints to study the effects of dietary interventions in small animal models of human metabolism or limited human clinical trials. Fundamental to the success of this approach is the ability to maintain large data sets and analyze complex multidimensional interactions (Ali et al., 2015).

17.2 NATURE OF GENETIC VARIATION

17.2.1 Genetic Epidemiology

Genetic epidemiology is "a science which deals with the etiology, distribution, and control of disease in groups of relatives and with inherited causes of disease in populations" (VanderWeele et al., 2014).

This field has provided initial proof of principle that genotypes do profoundly influence responses to diets or to nutrients in humans (Konstantinidou et al., 2014). A large number of studies have related variants in individual genes to disease risk, with varying degrees of success. However, it will be important that future studies consider more than just individual gene variants, one at a time. The genetic basis of many complex diseases appears to be a combination of relatively small variations in the DNA, typically in the form of SNPs or CNVs that increase disease susceptibility, rather than directly causing disease. For example, many of the genes associated with susceptibility to asthma and allergic response increase the risk of developing the disease by around 1.2-fold. However, carrying more than one such variant increases the risk significantly more. The complex group of genes associated with the major histocompatibility complex, and the role that this gene complex plays in IBD, provide a good example here (Goyette et al., 2015). Genetic epidemiology considers inherited causes of disease by looking at inheritance patterns, identifying the causal genes, and establishing which variants are associated with disease susceptibility (Abramson et al., 2014). This field also provides a necessary first step to studying gene–gene and gene–environment interactions, including nutrigenomics (Fig. 17.1).

Familial studies have traditionally been used for linkage analysis, but with the increased sensitivity of methods for analyzing genetic variation, such as high-density SNP microarrays, population studies are being increasingly used to establish associations between genes and disease risk. Next-generation sequencing (NGS) approaches also represent an increasingly promising option.

There has been rapid development of several NGS platforms, including Illumina, the Applied Biosystems SOLiD System, 454 Life Sciences (Roche), Helicos HeliScope, Complete Genomics, Pacific Biosciences PacBio, and Life Technologies Ion Torrent (reviewed in [Metzker, 2010]). There are also suppliers working on third-generation (sometimes referred to as next-next-gen) sequencing methods based on nanopore nucleic acid sequencing as well as single-molecule, real-time sequencing technology, which enables native DNA to be sequenced without cloning or amplification and with longer read lengths, typically exceeding 5 kilobases (kb) (Chaisson et al., 2015). All of these enable the relatively inexpensive production of large volumes of sequence data.

NGS approaches have been applied in the context of familial studies (Wang et al., 2015a), which, although still limited in terms of the number of participants due to cost, represent a great potential for a more comprehensive dissection of disease heritability in genomic epidemiology (Wang et al., 2015b). Regardless of the method, the key steps involve consideration of the following questions:

- Is there a genetic component to the disorder?
- What is the relative size of that component in relation to other causes of disease susceptibility?
- Which gene(s) are associated with the effect?
- Which are the functional variants in those genes?
- How will those variants interact with diet and/or other environmental factors to cause the phenotype?

Linkage studies consider the segregation of genes among family members and depend upon the cosegregation of two loci (one of which is the disease locus). They are dependent upon having access to a number of first-degree relatives for analysis. While association studies may also be family-based (transmission/disequilibrium test), they are increasingly being applied to population databases in genome-wide association studies (GWAS). For example, the Wellcome Trust Case Control Consortium involved 17,000 subjects, and follow-up studies also examined high numbers of individuals from different populations (Wellcome Trust Case Control, 2007). Such studies initially identified approximately seven variants associated with CD. Metaanalysis brought this number up to 71 (Franke et al., 2010). However, subsequent studies have used a higher density SNP chip. For example, the Immunochip, studying more than 75,000 cases and controls, identified a total of 163 IBD loci that met genome-wide significance (Jostins et al., 2012). Further high-density mapping brings this number significantly higher and has revealed key pathways involved (Goyette et al., 2015).

Twin studies provide important evidence on the genetic basis of disease by comparing monozygotic (MZ) twins (who share all of their genetic variants) with dizygotic (DZ) twins (who typically share half of their genetic variants). If there is evidence of a higher concordance between MZ as compared with DZ twins, this implies a genetic basis for the disease in question (MacGregor et al., 2000). Disease concordance of less than 100% between MZ twins implies an environmental component, exactly as

FIGURE 17.1 Depiction of how genetic epidemiology studies (top), particularly genome-wide association studies, can form the basis of nutrigenomics studies (bottom). *Text boxes* provide a description of the workflow of both study types. *SNP*, single nucleotide polymorphism.

seen in nutrigenomics. Such evidence has been used to suggest gene–environment interactions in the etiology of many diseases, including cancer (Lu et al., 2014), celiac disease (Greco et al., 2002), and schizophrenia (Xu et al., 2013).

Once there is evidence that the disease has a genetic basis, it becomes necessary to establish which genes are important. Much of the traditional work in this area has involved family studies, which consider allele frequencies in closely related subjects. Most powerfully, these involve multigenerational studies of large family pedigrees, but even parent/child trios can provide useful information. Segregation analysis can be used in order to establish Mendelian inheritance patterns or nonclassical patterns of inheritance, including mitochondrial diseases and genomic imprinting. These data can be subjected to linkage analysis in order to localize the crude chromosomal regions likely to contain the genes relevant to the disease. For example, the Online Mendelian Inheritance in Man (OMIM) database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim) typically identifies chromosome regions, based on linkage analysis. This yields broad chromosome regions harboring many genes, where finer resolution comes from recombination events (meiosis) in families. This approach has often

been used in the past since it does not require many markers, but it has historically been poor in terms of finding specific variants. However, NGS-based approaches now mean that familial studies can be carried out, which include fine mapping, and are therefore able to identify specific variants (Wang et al., 2015b). It is important that linkage studies are not confused with association studies, since they consider different questions. A comparison of these two approaches is provided in Table 17.1.

While linkage studies rely on recombinational analyses, association studies are based on linkage disequilibrium. This assumes that the SNP or CNV being studied is sufficiently close to the disease gene to result in an allelic association within the population (Austin et al., 2012; VanderWeele et al., 2014). CD must be considered as a successful example of a complex disease in which both linkage and association studies have been used to locate and find disease susceptibility genes (VanderWeele et al., 2014). Nevertheless, these studies are subject to the same pitfalls as conventional epidemiology, including selection bias, information bias, and confounding (Faye and Bull, 2011)

17.2.2 Single Nucleotide Polymorphisms

17.2.2.1 Single Nucleotide Polymorphism Discovery

Fundamental to the success of genetic epidemiology is the identification of novel associations between SNPs in certain genes and a particular phenotype. Whether established by GWAS or identified by decades' worth of linkage and candidate gene studies, large amounts of information on the genetic basis of many diseases is readily accessible through databases such as OMIM.

The increasing density and falling costs of DNA microarrays, and more recently of sequence-based approaches, has made GWAS the method of choice for SNP discovery. A number of microarray platforms allowing GWAS are available, but the field is largely dominated by solutions provided by Affymetrix and Illumina (see also Chapter 19). Affymetrix SNP chips use photolithography to synthesize 25-mer oligonucleotide probes in situ onto silicon wafers, with the Human SNP Array 6.0 having probes for >900,000 SNPs and a similar number of CNVs. Assays are performed by the hybridization of biotinylated cRNA, followed by washing, staining, and fluorescent detection. Illumina (https://www.illumina.com/) whole-genome genotyping assays use silica microbeads, each uniquely labeled with a 50-mer oligonucleotide probe. The Infinium HumanCore BeadChip array family, containing common variant tag SNPs, additional markers for sample tracking and quality control, and HumanCoreExome BeadChips, enabling extended coverage of exonic regions, provides up to 5,000,000 markers that are informative for SNPs and CNVs as well as mitochondrial DNA. Assays are performed by the hybridization of unlabeled genomic DNA, followed by single base extension with labeled nucleotides, washing, and fluorescent detection. High-throughput sample processing is enabled by iScan, HiScan, or HiScanSQ scanners.

Irrespective of the platform used, array-based GWAS provide a powerful new mechanism for the unselected investigation of disease relationships. Furthermore, an enormous amount of effort is also being put into bringing the cost of whole-genome sequencing down into the range of US $1000 (Ferguson et al., 2016) (see also Chapter 30).

NGS is generally not yet considered to be viable for large-scale whole-genome approaches, with factors such as library preparation for large numbers of samples being prohibitively expensive (Jin et al., 2016). However, it can be used in a more targeted fashion to identify disease-associated genetic variants. This can be applied for genes known to be associated with disease or combined with complementary genome-wide approaches. For example,

TABLE 17.1 Comparison of Linkage Versus Association Studies

Linkage Studies	Association Studies
Family-based	Families or unrelated individuals
Matching/ethnicity generally unimportant	Matching/ethnicity crucial
Few markers for genome coverage (e.g., 300–400 short tandem repeats)	Many markers required for genome coverage (10^5–10^6 single nucleotide polymorphisms)
Good for initial detection; poor for fine-mapping	Poor for initial detection; good for fine-mapping
Powerful for rare variants	Powerful for common variants; rare variant identification generally impossible
Identifies chromosome regions rather than individual genes	Identifies individual genes

one study used a combination of the Roche-NimbleGen EZ SeqCapV3 capture system and the Illumina HiSeq 2500 Rapid Run sequencing platform to investigate 650 known disease-associated genes (including untranslated and flanking intronic regions). The diseases covered by this approach included cardiac disorders, intellectual disabilities, hearing loss, muscular dystrophy, and complex Mendelian disorders. This approach identified 12 novel disease-associated variants (Delio et al., 2015). Furthermore, approaches such as pooled-DNA sequencing can enable NGS approaches for the identification of low-frequency and rare variants, which may represent a significant fraction of heritability not included in custom genotyping arrays (Jin et al., 2016).

17.2.2.2 *Single Nucleotide Polymorphism Analysis for Nutrigenomics*

Because of the rapid proliferation of published GWAS for many common diseases, SNP discovery is likely to be less important in nutrigenomics than the ability to genotype a population for which dietary and phenotype information is available, allowing correlations to be made. Over 100 different genotyping methods have been described to date, several of which have been discussed in detail in earlier chapters of this book. The ideal genotyping assay should have the qualities listed in Table 17.2.

Appropriate methods exist for genotyping a small number of samples for a single SNP through to thousands of samples for thousands of SNPs or even higher. Four of the most common genotyping methods have been compared for cost, specificity, and run-time, covering the range from low- to medium-high throughput (Delvaux et al., 2015). These authors reported that Tetra-primer amplification-refractory mutation system-polymerase chain reaction was more cost-effective, with superior sensitivity to the commonly used restriction fragment length polymorphism (RFLP), quantitative (q) PCR, or direct sequencing methods.

Huebner and coworkers compared several methods for studying a triallelic SNP (Huebner et al., 2007). RFLP, allelic discrimination PCR (ADPCR), TaqMan® (Applied Biosystems), LightCycler melting curve analysis (Roche Applied Science), the MassARRAY® System with iPLEX® chemistry (Agena Bioscience, formerly Sequenom), and DNA sequencing were used to genotype the triallelic SNP G2677/T/A (rs2032582) in the MDR1 gene of 73 individuals. Triallelic SNPs are increasingly being reported in the genome (Haridan et al., 2015) but are often initially reported as being diallelic due to limitations in common genotyping methods. To test the various available genotyping techniques, assay designs for genotyping analysis were based on the assumption that there were only two alleles (G and T), with the ability to detect the rare third A allele being assessed (Huebner et al., 2007).

RFLP and LightCycler methodologies reported genotypes for all 73 individuals, whereas the other methods had missing genotype rates between 4.1% and 5.5% (Table 17.3). More importantly, however, genotyping error rates were high for RFLP, ADPCR, and TaqMan®, due largely to their inability to correctly call the third allele. If the A allele is removed from the calculations, then the genotyping error rates for these assays are greatly improved, particularly for TaqMan® assays. Furthermore, if the presence of the third allele was known in advance, the ADPCR and TaqMan® assays could be redesigned to successfully detect this allele. LightCycler, MassARRAY®, and sequencing methods were all capable of detecting the third allele without prior knowledge of its existence, although for LightCycler and MassARRAY® methods, this ability initially required the examination of the individual

TABLE 17.2 Attributes of a Desirable Single Nucleotide Polymorphism Genotyping Method for Nutrigenomics Studies

Attribute	Characteristics
Simple assay design	An assay should be easily and quickly developed from sequence information and require as few manual steps as possible.
High throughput	Newer methods require high numbers of single nucleotide polymorphisms to be assayed in (preferably) thousands, or even tens of thousands, of samples.
Robustness	The method should tolerate suboptimal DNA quality and minor variations in assay conditions such as reagent volumes, incubation periods, and temperatures.
Unambiguous genotype calling	Genotype calling should ideally be automated if large numbers of samples and/or genotypes are being assayed.
Affordability	When calculating the cost per genotype, the expense of equipment, reagents, consumables, and personnel time should be allowed for.
Reliability	The assay must produce accurate and reliable results.

TABLE 17.3 Comparison of Six Common Single Nucleotide Polymorphism Genotyping Approaches for Genotyping a Triallelic Single Nucleotide Polymorphism

Method	Missing Genotypes	Genotyping Errors (G/T/A Alleles)	Genotyping Errors (G/T Alleles)
RFLP	0	12.3	3.0
ADPCR	5.5	16.9	6.8
TaqMan®	4.1	11.1	1.6
Light cycler	0	10.6	9.5
Sequenom	5.5	0	0
Sequencing	5.5	1.5	1.5

Numbers represent percentages.
Adapted from Huebner, C., Petermann, I., Browning, B.L., Shelling, A.N., Ferguson, L.R., 2007. Triallelic single nucleotide polymorphisms and genotyping error in genetic epidemiology studies: MDR1 (ABCB1) G2677/T/A as an example. Cancer Epidemiol. Biomarkers Prev. 16 (6), 1185–1192.

spectra rather than relying on automated genotype calling. The LightCycler assay suffered from high genotyping error rates, irrespective of whether the A allele was included in the analysis, negating the low rate of missing genotypes and the ability to detect the third allele. In comparison, the MassARRAY® and sequencing approaches had much lower genotyping error rates. The observed genotype frequencies for the MassARRAY® assay most closely matched the true genotype frequency, demonstrating that while a high genotype calling rate is important, it is more important that the genotypes are being called correctly.

Since this study was reported, there have been several advances in technologies for assessing SNPs, including droplet digital PCR (ddPCR) applications, such as the Bio-Rad QX100 system, which combines state-of-the-art microfluidics technology with TaqMan-based PCR to achieve highly sensitive and specific target DNA quantification. In addition to SNP genotyping, this technique can be applied to CNVs, rare variant detection, and transcript quantification (Mazaika and Homsy, 2014).

NGS is considered to be a promising approach for the identification of triallelic SNPs, with one study using Exome capture and sequencing (Illumina HiSeq, 2000 platform) in 221 Chinese individuals to identify over 2000 triallelic sites (Cao et al., 2015).

Based partly on results from these studies, a combination of genotyping approaches would seem appropriate, with the method determined by the throughput required. RFLP remains the simplest way of assaying a small number of samples for a small number of SNPs. The equipment and most of the reagents required to perform these assays will almost certainly be preexisting in any molecular biology laboratory, making the only significant cost the restriction enzyme. However, RFLP assays require considerable manual input, and the cost per genotype is linear, restricting its use to studies requiring low numbers of genotypes. These might include initial or exploratory studies, but the method would be impractical for the large nutrigenomics studies required to associate genotype with diet. TaqMan® assays are PCR-based homogenous assays performed in 96- or 384-well plates with fluorescent detection in a reverse transcription polymerase chain reaction instrument. This allows much greater throughput than RFLP, making TaqMan® assays ideal where an individual or relatively small number of SNPs need to be genotyped in larger populations of ~100 to several thousands. Due to the availability of a range of kit sizes, the cost per genotype tends to go down with an increasing number of samples, although each SNP must be assayed individually, meaning the cost per genotype is linear with increasing SNPs.

Increasingly, it is desirable to genotype the large numbers of SNPs identified in GWAS in a study population where specific dietary information can also be collected. To achieve this at an acceptable cost per genotype, multiplexed assays become necessary. The Agena Bioscience (formerly Sequenom) iPlex Gold genotyping assay is based on multiplex PCR followed by a single base primer extension and allelic discrimination by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (MS). The Assay Design Suite online tool (http://agenabio.com/products/massarray-system/software/) is used to design PCR primers and extension primers for multiplex reactions, currently up to 40-plex. Primer pools are created, and PCR amplification of DNA samples to be genotyped is performed in a 384-well format before unincorporated nucleotides are deactivated by a shrimp alkaline phosphatase treatment. Primer extension is then carried out on a thermocycler using the iPlex enzyme, followed by the addition of ion exchange resin to remove ions that would otherwise interfere with MS detection. Reactions are then spotted onto SpectroCHIPs, which are small silicon wafers with an array of 384 spots preloaded with a MALDI matrix optimized for DNA analysis. Finally, assay definitions and plate definitions are set up in Typer 4.0 software and spectra acquired from the SpectroCHIPs in the MassARRAY® Analyzer, a compact bench top MALDI-TOF MS.

The acquired genotype data can be reviewed in the Typer 4.0 software before export for statistical analysis. In this way, the MassARRAY® platform is capable of generating high numbers of genotypes per day at a low cost per genotype.

17.2.3 Copy Number Variants

A CNV has been defined as a copy number change involving a DNA fragment that is approximately 1 kb or larger (Haraksingh and Snyder, 2013). Changes involving the insertion or deletion of transposable elements are specifically excluded from the definition (Hehir-Kwa et al., 2013). There is reason to believe that most humans carry considerably more genetic variations in the form of CNVs than had previously been suspected and that these may be a greater source of human variation than SNPs (Abel and Duncavage, 2013). It is of interest that CNVs appear to be enriched within genes that are important in molecular–environmental interactions, potentially influencing immune defense and disease resistance or susceptibility in humans (Barber et al., 2005).

Whole-genome-based analyses such as array comparative genomic hybridization (see also Chapter 15) or the high-density SNP genotyping arrays from Affymetrix and Illumina will identify variants and establish their location. Redon and coworkers used high-density 500 K SNP genotyping arrays to reveal approximately 1200 CNVs of varying sizes (approximately 1 kb to over 3 Mb) (Redon et al., 2006). Such initial studies are being validated through repeated analysis, involving larger numbers of patient samples and different populations, using a similar technology (Hehir-Kwa et al., 2013). However, other assays are necessary to test specific hypotheses. Quantitative PCR-based methods such as TaqMan® analysis may be employed, but multiplex ligation-dependent probe amplification (see also Chapters 15 and 16) is also increasingly used to identify and screen gene-based deletions and duplications (Golzio and Katsanis, 2013). However, Shen and coworkers pointed out that the use of DNA probes developed for polymorphic nucleotides is not necessary for CNV identification and detection, and SNP-based assay design may even hinder successful CNV detection (Shen et al., 2008). They designed and evaluated a high-density array that used nonpolymorphic oligonucleotide probes for CNV detection, effectively uncoupling CNV detection from SNP genotyping. They were able to detect nearly 200 CNVs, of which approximately half appeared to be novel. These were independently validated using quantitative PCR. More sophisticated techniques have become important since that time (Abel and Duncavage, 2013).

NGS techniques are applicable in the case of CNVs. In one example, a targeted NGS approach combined capture array (to capture the coding sequence of 2181 genes associated with 561 Mendelian diseases) and NGS to detect mutations within these regions. The approach was highly accurate (99.95%), and in addition to detecting one CNV it was also able to detect a range of other mutations, including 62 SNPs, 14 insertions and deletions (Indels), 1 microdeletion, and 2 microduplications of chromosomes (Liu et al., 2015).

Other novel methods are also being applied to the detection of CNVs. As already mentioned, ddPCR can be used for this application (Mazaika and Homsy, 2014), as can the NanoString Technologies nCounter® system. This system is a multiplexed digital counting method based on fluorescent barcodes, which was originally developed for gene expression analysis (Geiss et al., 2008). However, it has been modified such that it can be used to derive absolute copy numbers within an individual (Brahmachary et al., 2014).

17.3 NUTRITIONAL EPIDEMIOLOGY

A fundamental assumption of nutrigenomics is that multigenic traits can be modified by a combination of nutritional and other environmental factors. If the field is to be successfully applied, accurate nutritional assessment must be a starting point. Traditional methods of analysis are well documented, for example, Willett, (1995). However, current assessment tools, including food frequency questionnaires, diet diaries, or diet recall methods, are notoriously unreliable. Bingham and coworkers systematically compared 8 dietary assessment methods in a sample of 160 women (a simple 24-h recall, a structured 24-h recall with portion size assessments using photographs, 2 food-frequency questionnaires, a 7-day estimated record or open-ended food diary, a structured food-frequency [menu] record, and a structured food-frequency [menu] record with portion sizes assessed using photographs) against weighed dietary records (Bingham et al., 1994). These authors found that 7-day estimated records or food diaries gave the most accurate information for standard nutrients. However, a nutritional role is increasingly attributed to bioactive dietary components that are not associated with traditional nutrition. More than 25,000 bioactive food components have been identified. These may impact health through a variety of mechanisms (e.g., interactions with the epigenome), and their effects may be influenced by a wide range of factors (e.g., genetic variation of the consumer) (Joven et al., 2014). No food composition database provides sufficient information on the levels of these compounds in common foods that could be used to estimate their consumption in the diet. Accurate nutritional assessment represents a major challenge to the field.

Nutrition researchers require methods of assessing food intake that will measure what people actually eat, instead of

what they remember eating. New methods are under development (e.g., cell phones are being used to report what people are eating on a real-time basis, using the camera feature of the cell phone to take a picture of the meal). The Food4me consortium used a cell phone-informed online tool of dietary recall and assessment (Fallaize et al., 2014; Forster et al., 2014).

Traditional food frequency data is useful for determining a typical intake of macro- and micronutrients, but it gives no indication of what short-term effects they might have on phenotype. In the case of CD, where diet can result in rapid changes in phenotype, knowing what foods people know they can or cannot eat is the most valuable information of all. Thus a qualitative measure of food tolerance can be used in addition to more conventional methods (Triggs et al., 2010).

17.4 EXPERIMENTAL MODELS

17.4.1 High-Throughput Assays

While genomic technologies make it possible to simultaneously examine the expression of tens of thousands of genes or detect hundreds of thousands of SNPs, and metabolomics is moving toward the identification of the multitude of compounds present in foods (see later), the tools to test the ability of foods to modulate phenotype against a specific genotype are also under continuous development.

One approach that shows promise is the in vitro, time-resolved fluorescence (TR-FRET) assay. In one study, 49 food extracts from a wide variety of food types (including grains, meat, fish, shellfish, dairy products, herbs, mushrooms, hops, fruits, and vegetables) were tested using TR-FRET to identify foods that contain Janus Kinase 2 (JAK2) inhibitors (Martin et al., 2015). Variants of the *JAK2* gene, which encodes the JAK2 tyrosine kinase, are associated with both CD and ulcerative colitis (UC) (Zhang et al., 2014). Several fruits, such as blackberry, boysenberry, and feijoa (also known as pineapple guava or guavasteen), were potent inhibitors of JAK2, and all of these fruits contain ellagitannins, which are known inhibitors of kinases. Such assays represent an approach for very quickly screening large numbers of food compounds for their effects on specific genes.

In vitro approaches can also be used to assess the effects of foods or food compounds on intestinal barrier functions. Maintenance of the epithelial barrier is critical to prevent intestinal contents such as food antigens from entering the body and causing an inflammatory response. Measuring *trans*-epithelial electrical resistance across Caco-2 cell monolayers (as a model of the intestinal epithelium) is a proxy for how "leaky" the epithelium is, and this approach has been used to assess the effects of various probiotics in maintaining barrier integrity (Anderson et al. 2010a,b). Such an approach can also be applied to assessing the effects of food compounds.

Another important factor to consider when investigating factors linked with intestinal inflammation is the role of the resident intestinal bacterial population, collectively referred to as the microbiota. The intestinal microbiota has been likened to an organ system (Marchesi et al., 2016), which carries out a wide range of functions that are indispensable for the human host. There is an increasing body of evidence that alterations in the microbiota may be associated with not only intestinal inflammation (Druart et al., 2014; Peterson et al., 2015) but also a range of other health outcomes, including obesity (Shen et al., 2014; Anhe et al., 2015). While investigating the microbiota in its entirety requires in vivo approaches, in vitro models are being developed, which can assess the interactions between specific microbial species and the intestinal epithelial cells. An example is a novel system, which enables the coculture of live obligate anaerobes with the (aerobic) human intestinal cell line Caco-2 (Ulluwishewa et al., 2015). In addition to investigating the interactions between microbe and host, such a model can enable food compounds to be added to the culture medium and thus provide a better understanding of the complex food—microbe—host interactions that may be critical for the maintenance of intestinal health.

17.4.2 Animal Models

Because of the complex interactions between foods, microflora, and host cells that occur within the intestine, in vitro models are limited in their ability to completely understand these interactions. While approaches such as the coculture system described earlier (Ulluwishewa et al., 2015) represent a significant step, even the most complex in vitro models will never contain every different cell type present in the intestinal or other tissues. Therefore in vivo models are required to gain a better understanding of changes in phenotype from healthy to disease states and the effects of foods and food components on this process.

A large number of animal models are available for specific purposes, conveniently categorized as:

- induced (experimental) disease models;
- spontaneous (mutant) disease models;
- genetically modified disease models;
  - transgenic;
  - knockouts; and
  - chemically induced.

Such models are available for all of the most common human diseases. In the following example, one specific group of such models is considered for colonic inflammation.

The authors' studies aim to understand how different foods or food components might interact with a particular genotype to cause the chronic intestinal inflammation that is a hallmark of CD. A systematic approach is needed, where intestinal parameters are measured at several time points, during which homeostasis of the intestinal system is challenged. The impacts on the biological gene and protein networks are measured, and physiology modeling systems biology approaches are used to integrate the key drivers of the effects of foods on intestinal wellness (Fig. 17.2).

Many different rodent models of intestinal inflammation could have been selected, with one comprehensive review identifying more than 69 spontaneous and induced rodent models that have been used for the study of either CD or UC (Neurath, 2012). This includes inducible colitis models and gene knockout or transgenic mouse and rat models. Although there is no one animal model that replicates every facet of human IBD (Prattis and Jurjus, 2015), all animal models of intestinal disease should exhibit some key disease attributes (similarity in morphology changes, inflammation status, pathophysiology, and time course). Studies with which the authors are involved are specifically looking at nutrient−gene interactions, using mouse models in which a well-defined genetic variation is associated with intestinal inflammation. At least 20 rodent models of IBD involving single gene manipulations exist, with some of the symptoms characteristic of IBD (Valatas et al., 2013). Two have proven to be particularly useful.

17.4.2.1 Interleukin 10 Gene Deficient (Il10$^{-/-}$) Mouse

IL10 is an antiinflammatory cytokine, and a defect in the IL10 gene leads to an imbalance in the inflammatory responses. While *Il10*$^{-/-}$ mice (C57BL/6 J/129-Ola background) have been reported to develop Crohn's-like colitis by 12 weeks of age when raised under conventional conditions (Berg et al., 1996), the level of inflammation observed is influenced by the background strain. In the case of the C57BL/6J background, more consistent inflammation is observed when the *Il10*$^{-/-}$ mice are inoculated with *Enterococcus faecalis* (Balish and Warner, 2002) or *Enterococcus* isolates (both *faecalis* and *faecium*) in combination with conventional intestinal flora derived from healthy C57BL/6J mice raised under conventional conditions (Roy et al., 2007; Barnett et al., 2010). As is the case in human IBD, the precise mechanism that results in inflammation in *Il10*$^{-/-}$ mice is unclear, although evidence suggests an inappropriate inflammatory response to normal intestinal flora through activation of the $CD4^{+}$ th1 cells and the depletion of their inhibitors, that is, the regulatory T cells (Kuhn et al., 1993).

Although *Il10*$^{-/-}$ mice have IBD-like symptoms, variants in this gene were not initially associated with IBD in human populations (Klein et al., 2000; Castro-Santos et al., 2006). However, SNPs near the 3′ UTR of the *IL10* gene were associated with UC (Franke et al., 2008), and variants of the IL10 receptor gene have been linked to early-onset IBD (Beser et al., 2015). In addition, variants in the IL23 receptor have shown an association with IBD in a number of population groups, including those from New Zealand (Roberts et al., 2007), the Netherlands (Weersma et al., 2008), and Finland (Lappalainen et al., 2008). It has been shown that the T cell-mediated colitis that develops in *Il10*$^{-/-}$ mice may be dependent on IL23 (Yen et al., 2006). This mouse model has been used by the authors and others to identify food components that may ameliorate or prevent human IBD, including polyunsaturated fatty acids (Knoch et al., 2009; Knoch et al., 2010b; Cooney et al., 2012), fish oils (Hegazi et al., 2006), flaxseed oil (Cohen et al., 2005), grape seed extract (Wang et al., 2013; Yang et al., 2015), and some probiotics (McCarthy et al., 2003; Shi et al., 2014).

*17.4.2.2 Multidrug Resistance Gene Deficient (*Mdr1a*) Mouse*

The Mdr1a mouse carries a disruption of the multi-drug resistance gene *Abcb1a* (a.k.a. Mdr1a), which encodes the transporter P-glycoprotein 3, or MDR1. MDR1 belongs to a family of transmembrane transporters, known as ATP-binding cassette transporters, and a number of mutations that lead to reduced MDR1 activity have been linked to IBD (Brant et al., 2003; Ho et al., 2006). In the intestinal tract, MDR1 is expressed on the luminal surface of epithelial cells, pumping drugs and toxins (e.g., those produced by certain strains of bacteria) from the cell membrane and cytoplasm back into the intestinal lumen (Ho et al., 2003; Bilsborough and Viney, 2004); thus a failure to "clear" these toxic compounds may be triggering and maintaining the spontaneous intestinal inflammation observed in Mdr1a mice (Panwala et al., 1998; Banner et al., 2004; Wilk et al., 2005; Dommels et al., 2007). Furthermore, antibiotic treatment has been shown to prevent and therapeutically reverse inflammation in Mdr1a mice, demonstrating a role for bacterial flora in the initiation and progression of inflammation in these mice (Panwala et al., 1998) and the importance of considering the bacterial flora in any gut-health studies. The Mdr1a mouse model is another tool the authors have used to understand the development of intestinal inflammation (Dommels et al., 2007) and to test the efficacy of food components such as curcumin and rutin (Nones et al., 2009; Cooney et al., 2016) and green tea (Barnett et al., 2013) on the prevention of this phenotype.

For gene-mutation IBD rodent models as well as monitoring the effect of food components on the physical and biochemical signs of inflammation, gene (Knoch et al.,

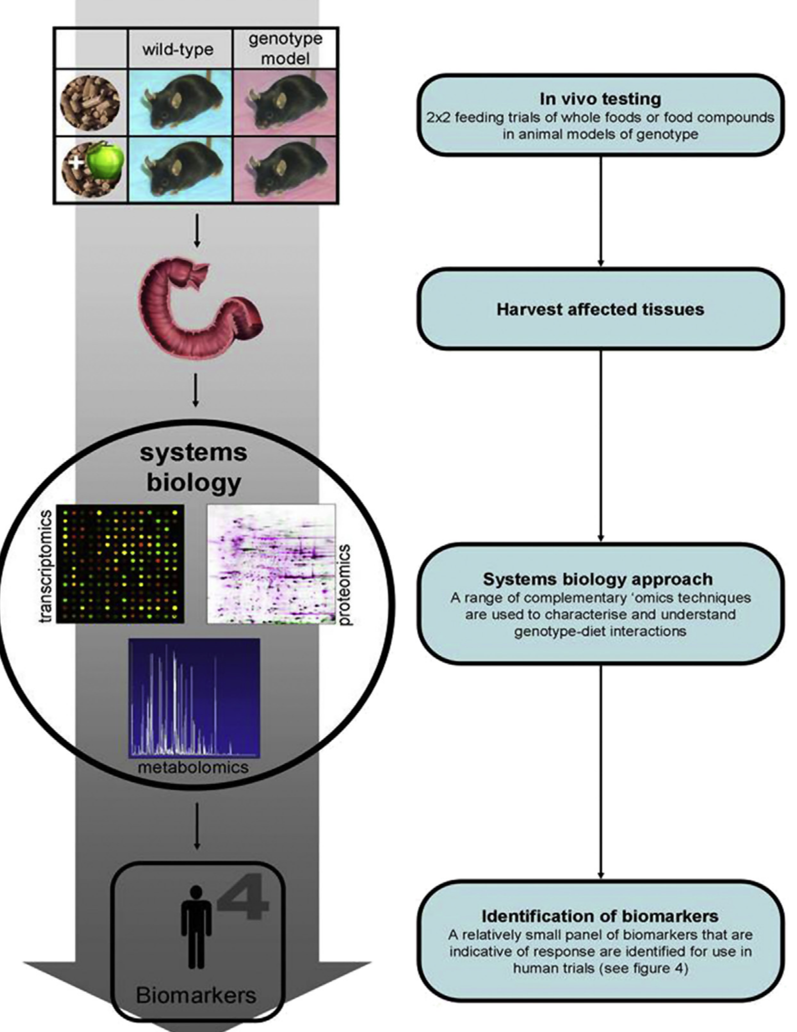


FIGURE 17.2 Systems biology approach employed for animal model studies to understand how different foods or food compounds might interact with a particular genotype. One output of this approach is a defined set of biomarkers for use in human trials (see also Fig. 17.4).

2009; Nones et al., 2009) and protein (Cooney et al., 2012; Barnett et al., 2013) expression, metabolite profiling (Lin et al., 2010; Lin et al., 2011), and epigenetic changes such as DNA methylation (Denizot et al., 2015) and chromatin binding (Chahar et al., 2014) in the colon cells can be assessed to understand the mechanisms by which changes in inflammation occur. Changes in the expression of genes implicated in human IBD, for example, those involved in processes such as immune response, inflammation, antigen presentation, and xenobiotic metabolism (Dommels et al., 2007; Barnett et al., 2010), are of particular interest. This is the case in terms of the inflammation per se, that is, establishing or confirming the relevance of a particular model (Dommels et al., 2007; Roy et al., 2007; Barnett et al., 2010; Knoch et al., 2010a), and in providing insight into how particular foods or food components might ameliorate symptoms linked to human IBD (Knoch et al., 2009; Knoch et al., 2010b; Cooney et al., 2012; Paturi et al., 2012; Barnett et al., 2013).

17.4.3 Human Clinical Trials

The ultimate proof that a nutritional intervention is beneficial to health utilizes a double-blind, placebo-controlled clinical trial. Optimal study design would utilize randomly chosen subjects who are prospectively examined and compared to well-matched control individuals using genomically-based diagnostic techniques. However, such studies are not cost-effective in terms of either time or money. Furthermore, an increasing number of such trials have provided negative or ambiguous results, even where there is a very strong reason to believe that the foods in question are beneficial. Examples

would include omega-3 polyunsaturated fatty acids, where the metaanalysis of several large international studies has failed to support a benefit against cardiovascular disease (Hooper et al., 2006). Similarly, despite good theoretical grounds for the hypothesis, most human intervention studies on antioxidants have given disappointing results (Moller and Loft, 2004, 2006). More generally, most human clinical studies on randomly selected populations have failed to demonstrate predicted outcomes from dietary interventions alone. One possible explanation is that many of the studies have merely provided a better description of homeostasis, showing the robustness of health and the adaptability of the human body while not including the appropriate methodology to assess the more analytically elusive subtleties of multiple minor changes. Two approaches to address this lack of results from clinical studies are yielding fruit and should not be considered to be mutually exclusive.

17.4.3.1 Subject Preselection or Stratification According to Genotype

A flow diagram showing how a trial involving subject preselection or stratification according to genotype might optimally work is provided in Fig. 17.3.

For example, a study by Hallman and coworkers reported that they could predict fatty acid profiles in blood from a combination of dietary fat intake and genotyping of the FADS1 rs174546 SNP in 1607 participants from the Food4me study (Hallmann et al., 2015). They were seeking to improve the ratio of omega-3 to omega-6 polyunsaturated fatty acids in the blood, and this was optimized by combining the dietary intake data with the genotype. Similarly, a metaanalysis from this same consortium validated the association between the fat mass and obesity-associated gene, obesity, and reported food intake (Livingstone et al., 2015).

17.4.3.2 Challenging Homeostasis

Ultimately, the goal of nutrigenomics is to ensure optimum health. However, the quantification of health cannot be achieved by applying "absence of disease" approaches. Intermediate processes such as metabolic stress, oxidative stress, or inflammatory stress may be more amenable to biomarker approaches. However, these processes strongly influence each other, and their complexity means that there is no completely satisfactory simple biomarker. Interindividual variation may be larger than the effect imposed by the treatment, partly because the "confounders" cannot be adequately controlled. Genetics, lifestyle, age, stress, concurrent infections, etc. all determine the health status of an individual and need to be quantified.

For these reasons, newer clinical trial designs look to challenging homeostasis. The concept considers that only after stressing or perturbing homeostasis can the robustness ("health") of the system be demonstrated and potentially quantified. A classic example in nutrition research is the oral glucose tolerance test, whereby individuals who have fasted overnight are provided with a high-glucose bolus, and the level of glucose in their bloodstream is measured immediately and again after 2 h. Another example being studied by the Nutritech consortium is a high-fat challenge test, where individuals in different states of metabolic health are compared for "phenotypic flexibility" (Stroeve et al., 2015). This concept implies that health can be measured by the ability to adapt to a temporary condition of stress.

17.5 DEFINING THE PHENOTYPE

The term "deep phenotyping" means "…more complete descriptions of the physical state of individuals (and, by aggregation, groups of individuals) than has been possible in the past…" (Baynam et al., 2015). The justifications for this are 3-fold:

- Translational research is necessary for the translation of dietary recommendations but is insufficient in providing a definitive proof of efficacy.
- Phenotype variability dilutes statistical power and the strength of association.
- There are differences in pathophysiology over time within an individual.

A good example is provided by the field of cardiovascular disease, in which three phases can be identified, whereby it has become increasingly possible to define a serum–lipid profile that differentiates healthy individuals from those at a significantly increased risk of developing disease (Fig. 17.4). This modified lipid profile that integrates not only serum cholesterol but also HDL and LDL cholesterol provides a basis whereby the effects of a dietary intervention can be readily measured. The aim is to move individuals from the "high-risk" to the "low-risk" profile, with good predictive powers for benefit or otherwise.

More powerfully, deep phenotyping involves the application of "omics" technologies to better understand the underlying biology and metabolism associated with disease, thereby enhancing diagnosis and potentially identifying improved phenotypic markers of disease state. International standards are being applied to the development of guidelines and standards for these technologies. For example, *Nature Biotechnology* has an online community consultation initiative (http://www.nature.com/nbt/consult/index.html) that is intended to encourage researchers to participate across country barriers.

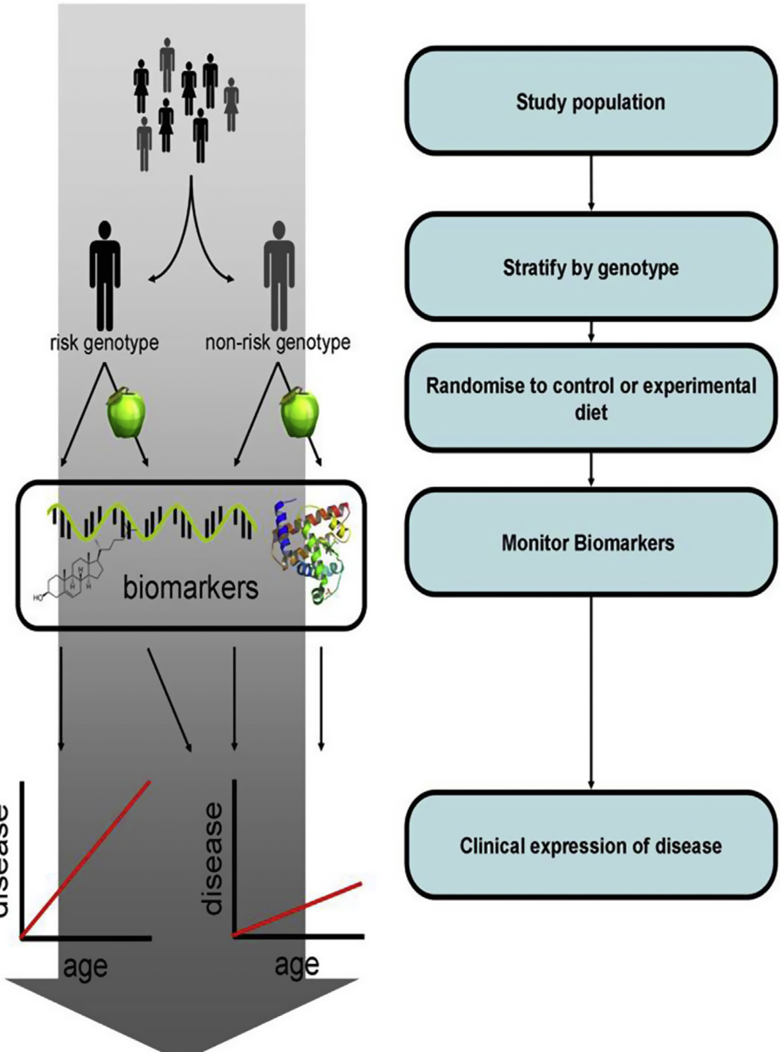


FIGURE 17.3 Study design for human trials of foods where subjects are preselected or stratified according to genotype. Study participants are randomized to receive either a control diet or an experimental diet that includes the food shown to ameliorate the phenotype associated with a particular genotype in animal models (see also Fig. 17.2). Biomarkers previously identified in the same animal studies are monitored to rapidly determine if the experimental diet can restore the nonrisk-genotype profile without the need to follow study participants until the development of disease. These results can be extrapolated to predict long-term protection against disease development.

Because deep phenotyping relies on a suite of new and emerging technologies, the costs can be considerable. For example, one cholesterol measurement might cost US $5−10, whereas an appropriately planned and performed microarray-based expression experiment could cost in the order of US $500 per sample. Specialized tissue collection and complex bioinformatic analysis only add to this cost. However, the end result of preliminary analyses using these methods may be a relatively small number of measurements (be they gene or protein expression events or blood/urine metabolites), which, when combined, give a very accurate and precise diagnosis of a disease. Most importantly, such measures may be possible before there is any obvious disease onset, providing the ability to prevent or treat symptoms and therefore improving the prognosis for the individual in question. Techniques that are becoming routinely applicable to nutrigenomics research hold much promise in the field of disease diagnosis.

17.5.1 Transcriptomics

The transcriptome is the set of all messenger RNA (mRNA) molecules, or "transcripts," produced in one or a population of cells (Duffus et al., 2007). The transcriptome reflects the genes that are being actively expressed at any given time because it includes all *mRNA* transcripts in the cell. Transcriptome analyses have used high-throughput techniques based on DNA microarrays, although increasingly sequence-based approaches are being applied.

While originally developed using arrays of cDNA "spotted" onto glass microscope slides, the field of transcriptomics (as in

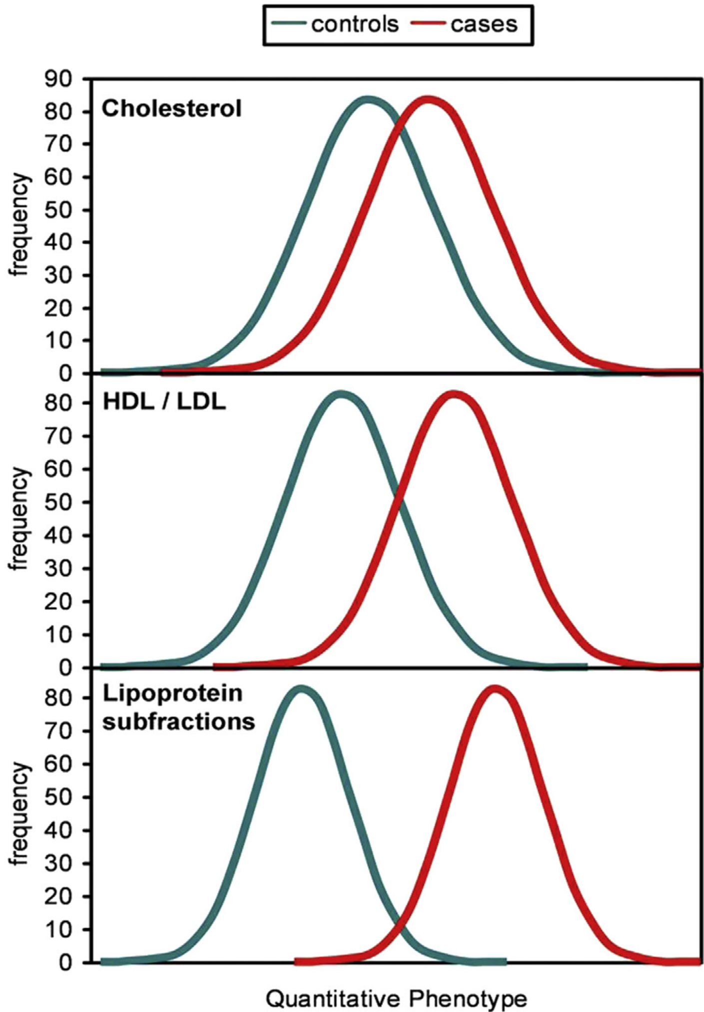


FIGURE 17.4 Example of how deep phenotyping can be used to better define disease risk. Since the 1950s it has been recognized that cholesterol levels are associated with an increased risk of cardiovascular disease but that the predictive power of cholesterol levels was relatively poor (top panel). In the 1980s it was demonstrated that a wider lipid profile investigating HDL and LDL was an improved predictor of cardiovascular disease (middle panel). More recently, a detailed examination of lipoprotein subfractions has been used to establish much more accurate cardiovascular disease risk predictions (lower panel). Dietary intervention can then be used to alter an individual's lipid profile to shift him or her from the high-risk to the low-risk profile.

the case of GWAS) is now dominated by microarray solutions provided by companies such as Affymetrix, Agilent Technologies, and Illumina. As mentioned previously, Affymetrix arrays, like the SNP chips, consist of in situ synthesized 25-mer oligonucleotide probes applied using photolithography technology. Agilent Technologies also use in situ hybridized oligos, but these are synthesized using piezo-electric (ink-jet) technology and have a length of 60 bases. Illumina BeadChips use 50-mer oligonucleotides immobilized onto microbeads, which then self-assemble onto ordered microwell-etched substrates. A summary of the key features of these array technologies can be found in Table 17.4.

TABLE 17.4 Features of Three Currently Available Expression Array Technologies

Supplier	Affymetrix	Agilent	Illumina
Feature	Oligonucleotide	Oligonucleotide	Oligonucleotide
Length	25-mer	60-mer	50-mer
Color	One-color	One- or two-color	One-color
Number of unique probes	~5,500,000 (x1 sample/chip)	~44,000 (x8 samples/slide)	~48,000 (x12 samples/slide, 15-fold redundancy)

The use of RNA-Seq (i.e., transcriptome profiling using NGS) is increasingly being applied for transcriptomics studies. RNA-Seq is considered to be superior to microarrays, with advantages such as a better detection of low-abundance transcripts, enabling detection of genetic variants, and having a broader dynamic range than microarrays (Zhao et al., 2014). However, while the cost is rapidly decreasing, RNA-Seq is still more expensive than comparable array-based technologies, and the relatively large amount of data it yields presents a challenge both in terms of storage and analysis. For this reason microarrays are still more commonly used (Zhao et al., 2014). As is the case for other NGS applications, a range of platforms can be used for RNA-Seq analysis.

For a genome-wide measurement of gene expression changes associated with intestinal inflammation in both $Il10^{-/-}$ (Barnett et al., 2010) and Mdr1a mice (Dommels et al., 2007), Nutrigenomics New Zealand has used Agilent Technologies microarrays. Briefly, cyanine-3- and cyanine-5-labelled cRNA is synthesized from purified colon RNA and hybridized to Agilent mouse whole-genome arrays. The slides are scanned using an Agilent Technologies microarray scanner. The digitized data are aligned, and the spot intensity data are downloaded as text files using Feature Extraction software. Linear models for microarray analysis (Bioconductor) (Smyth, 2005) are then used for the normalization and identification of differentially expressed mRNA with multiple comparisons to minimize false discoveries using internationally recognized standards (Cheng and Pounds, 2007). Bioconductor (http://www.bioconductor.org) is open source and open development software for the "analysis and comprehension of high-throughput genomic data."

The assessment of data from RNA-Seq experiments brings additional challenges. Whereas microarray probes are based on the existing knowledge of a reference sequence, this is not the case with all data from RNA-Seq. The simplest and most accurate approach involves the use of curated annotations, but it is also possible to use computational approaches, either genome-guided (in which sequence reads are aligned to a reference genome) or de novo (where the reads are assembled without the aid of reference data). As described in one comparative study (Janes et al., 2015), using curated annotations is superior to available computational approaches.

As is the case for microarrays, Bioconductor workflows are available for the analysis of RNA-Seq data, for example, the "rnaseqGene" package, which enables alignment to a reference genome, read counts, quality assessment, analysis of differential gene expression, and visual exploration of RNA-Seq data (Love et al., 2015).

Having identified differentially expressed mRNA probes, there are many approaches with which to interpret the biological significance of these changes, for example, understanding the biological pathways and/or functions to which they contribute. We have utilized applications such as Ingenuity Pathways Analysis (IPA; http://www.ingenuity.com/) for the analysis of transcriptome data from a variety of studies (Knoch et al., 2009; Barnett et al., 2010). IPA, now part of the Qiagen "family," is a web-based application that uses a curated knowledge base of information to enable the interpretation of a range of omics data, including microarray and RNA-Seq transcriptome data. Because of the proprietary knowledge base on which it relies, the use of IPA software represents a significant cost. However, there are also many free analysis packages that are available, such as DAVID (http://david.abcc.ncifcrf.gov/) for analyzing candidate gene lists; PathVisio (www.pathvisio.org; discussed in more detail later with respect to integrating omics data sets) for pathway statistics and data visualization; and Cytoscape (www.cytoscape.org) for visualizing and integrating complex networks. The authors have also used approaches such as gene set enrichment analysis (performed using Bioconductor), which, unlike gene-centric analyses, takes into account the expression of a group of genes (e.g., those in a common pathway) as a whole, rather than examining individual genes (Russ et al., 2013). This allows for the fact that alterations in the activity of biological pathways may be manifested by small but consistent changes in the expression of multiple genes within that pathway.

17.5.2 Epigenomics

Epigenetic mechanisms, including DNA methylation, histone modifications, and noncoding RNAs, are increasingly recognized as important in the regulation of cellular processes in response to environmental stimuli such as diet. Epigenomics is the analysis of these mechanisms on a genome-wide level. While the genome provides information about the potential nature of the cellular proteins, and the transcriptome can show which of these proteins are currently being transcribed via mRNA, epigenetic mechanisms represent a key point at which the expression of genes can be regulated.

A range of techniques is available for the assessment of epigenetic changes. As is the case with genome and transcriptome studies, many of these are moving toward the assessment of genome-wide analyses, such as methylation-specific amplification microarray (MSAM) to screen for changes in methylation status across the genome (Kellermayer et al., 2010) or ChIP-on-chip to determine changes in histone modifications at a genome-wide scale. Not surprisingly, as the use of genome-wide approaches increases, so the use of NGS technologies is also becoming increasingly important in the assessment of changes in the epigenome. For DNA methylation this can include such methods as MSA-Seq, in which the methylation-specific amplification used

for MSAM is adapted for NGS techniques. In the case of histone modifications, ChIP-Seq is used rather than array-based approaches, although as is the case for RNA-Seq, ChIP-on-chip is still an attractive option due to the relatively lower cost and ease of analysis (Yoder, 2015). NGS can be relatively easily applied to the assessment of non-coding RNA, in some cases enriching the sample for small RNA species such as miRNAs prior to sequencing. As is the case for all high-throughput approaches, the biggest challenge is in the analysis of the large quantities of data that are available and in linking these to the other omics data sets.

17.5.3 Proteomics

Proteomics is the analysis of the protein complement present in a cell, organ, or organism at any given time (Baynam et al., 2015; see also Chapter 21). While the genome provides information about the potential nature of the cellular proteins, the actual protein composition ("the proteome") ultimately determines the phenotype. Fundamental technologies for the separation of proteins and/or peptides are one- and two-dimensional gel electrophoresis and one- and two-dimensional liquid chromatography, typically coupled with MS (see also Chapter 24). Baynam et al. (2015) provide a comprehensive description of current technologies for proteome fractionation and the separation of biological samples.

In order to allow direct comparisons with gene expression measurements, the same tissue sample can be extracted to provide a source for both RNA and protein. This is possible using TriZOL reagent or kits such as the AllPrep DNA/RNA/Protein Mini Kit (Qiagen), which purifies RNA, DNA, and protein from the same cell or tissue sample.

The model of intestinal inflammation again provides an example of workflow in proteomics. The proteomic method that has been commonly employed within Nutrigenomics New Zealand is two-dimensional gel electrophoresis using differential in gel electrophoresis (DIGE) technology, by which proteins differentially expressed during intestinal inflammation can be recognized and subsequently identified using MS technology. Although multidimensional liquid chromatography-MS approaches are increasingly being applied for qualitative and quantitative proteome analysis (as reviewed in Di Palma et al., 2012 and Wu et al., 2012), a combined two-dimensional DIGE/MS approach is still widely used. For example, a PubMed search for "proteomics 2D DIGE" identified 160 publications in 2014, and we have published studies that have utilized this technique (Barnett et al., 2013; Cooney et al., 2016).

17.5.4 Metabolomics

Metabolomics is increasingly being used in human nutrition and applied in a wide range of studies, including IBD and other autoimmune diseases (De Preter and Verbeke, 2013; Ismail et al., 2013; Coughlin, 2014; Dawiskiba et al., 2014; Fiocchi, 2014; Kang et al., 2015). The ultimate goal is to understand the effects of dietary components on human metabolic regulation. Knowledge is accumulating as to how changes in the nutrient content of the human diet lead to changes in metabolic profiles. These are complex, because they result from at least three different sets of signals:

- nutrients and bioactive compounds that are absorbed from the diet,
- xenobiotics that are absorbed and metabolized from their environmental sources, and
- the large-bowel microflora, which also produces significant metabolic signals.

Metabolic profiling uses pattern-recognition statistics on assigned and unassigned metabolite signals and the collection of comprehensive data sets of identified metabolites. The technology has the potential to distinguish between different dietary treatments, which would not have been targeted with conventional techniques. The field faces significant technical challenges due to the sheer number of potential metabolites and the widely varying chemical and physical properties of them. Thus the ultimate success of this approach will be heavily dependent upon the development of libraries of small molecules to aid in metabolite identification. However, a key advantage to metabolomics is the potential for relatively noninvasive tissue collection (e.g., fecal, urine, blood, buccal swabs, or breath volatiles).

Two key technologies for metabolomics research are nuclear magnetic resonance (NMR)- and MS-based. For example, urine metabolites associated with inflammation in the Mdr1a mouse model have been identified using NMR analysis (Dryland et al., 2008), and gas choromatography—MS technology has been applied in the case of the *Il10*$^{-/-}$ model (Lin et al., 2010; Lin et al., 2011). It is anticipated that either of these approaches will be applicable to human samples.

17.5.5 Microbiota Analysis

As already mentioned, there is an increasing body of evidence that the population of microbes inhabiting the gastrointestinal tract, collectively known as the intestinal microbiota, plays a vital role in human health. This includes carrying out a range of metabolic processes that are essential to the host, such as the synthesis of vitamins, supporting the notion that the microbiota has been likened to an organ in its own right (Marchesi et al., 2016). Because of the essential role of the microbiota, it is clear that changes in the structure and/or function of these bacterial populations can have important implications for disease, including intestinal inflammation (Druart et al., 2014; Peterson et al., 2015) and obesity (Shen et al., 2014; Anhe et al., 2015).

The study of intestinal microbes has previously been limited due to the need for culture-based techniques, which are extremely difficult for the strictly anaerobic bacteria that comprise the vast majority of the microbiota. However, DNA-based, culture-independent techniques, and in particular NGS approaches, have revolutionized the study of the microbiota. Not only can the species composition of the microbiota be accurately assessed using 16s rRNA sequencing, but comprehensive sequencing of the collective microbial genome (the so-called "metagenome" or "microbiome") also enables the functional capacity of the microbiota to be assessed, while RNA-Seq approaches allow the expression of microbial genes to be assessed, providing unprecedented insight into the functional status of the microbiome at a given time. Due to the complex relationship between the host and microbial genomes, this information can be integrated with, for example, assessment of the host's gene expression profile to understand how the complex interplay between host and microbial genes may impact the health of the host.

17.5.6 Biomarkers

A biomarker provides a surrogate endpoint that indicates a probability of disease. The use of biomarkers to predict disease states (e.g., blood pressure for cardiovascular disease) has been effective in directing clinicians' recommendations for intervention and treatment. Similarly, biomarkers can be used as indicators of disease risk reduction following dietary manipulation (Bennike et al., 2014; Park et al., 2014). The importance of identifying biomarkers is to identify signs of disease susceptibility as early as possible, so that interventions have a good probability of impacting the progression of disease and improving outcomes. A desirable nutrigenomics research agenda should include the investigation of a set of changes that lead to a desirable health endpoint. These changes should be investigated in observational studies in carefully selected populations, with nutrient/dietary pattern exposure data calibrated using specific validated exposure biomarkers. This would be a substantial intervention development/initial testing research enterprise, including small-scale human feeding studies with biomarkers of disease risk as outcomes, before there is collaborative agreement among basic, clinical, and population scientists for assessing readiness for dietary intervention randomized controlled trials.

It is beyond the scope of this chapter to comprehensively cover biomarkers, but more details are provided online (http://biomarker.encode.net.nz/). While there is value in developing and validating biomarkers for specific disease states, such as cancer (Heckman-Stoddard and Smith, 2014), there may be more value in general biomarkers of disease processes such as inflammation (Nanau and Neuman, 2012; Serban and Dragan, 2014), since this is involved in a considerable number of different diseases. Important advances utilize genomic technologies for biomarker discovery. For example, there is evidence that the broad application of proteomic technologies has the potential to accelerate our understanding of the molecular mechanisms underlying disease and may facilitate the discovery of new drug targets and diagnostic disease markers (Park et al., 2014). These authors describe the current technologies for proteome fractionation and separation of biological samples, based on their own laboratory workflow for biomarker discovery and validation.

17.6 INTEGRATING COMPLEX DATA SETS: DATA MANAGEMENT, BIOINFORMATICS, AND STATISTICS

The model that Nutrigenomics New Zealand has selected for tailoring foods according to genotype requires an intelligent choice of food sources, fractionating these foods into class fractions and chemically characterizing food sources and food components for use in "high-throughput" nutrient sensor arrays. For example, initial screens suggested that several fruits, including blackberry, feijoa, and pomegranate, might be good sources of compounds to counter CD in specific populations (Martin et al., 2015). High-throughput screens have been used to generate data for a food-component efficacy database, which is integrated into the bioinformatics platform and then used to choose material for testing against other diet/genome targets appropriate to health, performance, and/or disease. Very large data sets are generated by these means.

A major initiative is necessary to deliver bioinformatics, biostatistics, and database resources for nutrigenomics studies. This typically involves the establishment of a relational database, which can be interrogated for relationships between genetic variation and disease states plus the impact of diet in human populations. It also requires the development of formal infrastructural platforms for bioinformatics and database resources in order to analyze microarray and proteomics data on gene expression patterns, identify clusters of affected genes, discover human analogues, and extrapolate this information to human health or disease. Reducing the dimensionality of these complex data sets remains a significant analytical challenge (Kaput and Dawson, 2007), and it has been noted that there is an increasing lag between the ability to generate "big" data and the ability to analyze these data sets effectively (Alyass et al., 2015).

A range of tools is being developed to address the challenges of dealing with large omics data sets. This includes packages such PathVisio, which was first released in 2008 (van Iersel et al., 2008) as a graphical editor for biological pathways with a particular focus on interpreting high-throughput data sets. PathVisio has undergone

continuous development since its first release, with numerous plug-ins developed and the ability to integrate with external databases such as WikiPathways. It has also been used to integrate data sets from different analyses, for example, mRNA and microRNA (Kutmon et al., 2015). Another useful tool is "mixOmics" (http://mixomics.org/introduction/), which is in the R free software environment (https://www.r-project.org/). mixOmics has a particular focus on graphical representations to study relationships between two "omics" data sets and better understand their complex associations (Gonzalez et al., 2012). One review summarized the available so-called "translational" platforms, which integrate clinical and omics data in the context of personalized medicine (Canuel et al., 2015). Such platforms must be able to store and integrate clinical and omics data and enable analysis and access to external databases as well as ensuring the privacy of individuals' data. These challenges clearly also need to be met in the case of personalized nutrition.

17.7 CONCLUSIONS

The Nutrigenomics New Zealand model provides an approach toward personalized, genotype-based nutrition that has the potential to provide food products and personalized advice to benefit health at the individual or population level. There is convincing evidence that SNPs in certain genes may profoundly influence the biological response to nutrients. However, the effects of single-gene variants on risk or risk factor levels of a complex disease tend to be small and inconsistent. Increased sensitivity of current biological measurements plus methods of integrating information on combinations of relevant SNPs or CNVs in different genes will become necessary to move the field to a higher dimension. Many of the challenges are in bioinformatics, especially in relation to reducing the complexity of multidimensional data sets (Coughlin, 2014; Fiocchi, 2014; Eijssen et al., 2015). To date, there are only sporadic examples of clinical trials utilizing these technologies, and we have not investigated the potential adverse effects of a genotype-derived dietary intervention. There are many issues to be addressed before genomic approaches become broadly accepted for guiding food development or nutritional recommendations.

ACKNOWLEDGMENTS

Nutrigenomics New Zealand is a collaboration between AgResearch Ltd., Plant & Food Research, and The University of Auckland, with funding through the New Zealand Ministry of Business, Innovation & Employment. The authors acknowledge the contribution of Dr. Martin Philpott to the previous edition of this chapter.

REFERENCES

Abel, H.J., Duncavage, E.J., 2013. Detection of structural DNA variation from next generation sequencing data: a review of informatic approaches. Cancer Genet. 206 (12), 432−440.

Abramson, M.J., Perret, J.L., Dharmage, S.C., McDonald, V.M., McDonald, C.F., 2014. Distinguishing adult-onset asthma from COPD: a review and a new approach. Int. J. Chronic. Obstruct. Pulm. Dis. 9, 945−962.

Ali, F., Ismail, A., Esa, N.M., Pei, C.P., 2015. Transcriptomics expression analysis to unveil the molecular mechanisms underlying the cocoa polyphenol treatment in diet-induced obesity rats. Genomics 105 (1), 23−30.

Alyass, A., Turcotte, M., Meyre, D., 2015. From big data analysis to personalized medicine for all: challenges and opportunities. BMC Med. Genomics 8, 33.

Anderson, R.C., Cookson, A.L., McNabb, W.C., Kelly, W.J., Roy, N.C., 2010a. *Lactobacillus plantarum* DSM 2648 is a potential probiotic that enhances intestinal barrier function. FEMS Microbiol. Lett. 309 (2), 184−192.

Anderson, R.C., Cookson, A.L., McNabb, W.C., Park, Z., McCann, M.J., Kelly, W.J., Roy, N.C., 2010b. *Lactobacillus plantarum* MB452 enhances the function of the intestinal barrier by increasing the expression levels of genes involved in tight junction formation. BMC Microbiol. 10, 316.

Anhe, F.F., Varin, T.V., Le Barz, M., Desjardins, Y., Levy, E., Roy, D., Marette, A., 2015. Gut microbiota dysbiosis in obesity-linked metabolic diseases and prebiotic potential of polyphenol-rich extracts. Curr. Obes. Rep. 4 (4), 389−400.

Austin, M.A., Hair, M.S., Fullerton, S.M., 2012. Research guidelines in the era of large-scale collaborations: an analysis of genome-wide Association Study Consortia. Am. J. Epidemiol. 175 (9), 962−969.

Balish, E., Warner, T., 2002. *Enterococcus faecalis* induces inflammatory bowel disease in interleukin-10 knockout mice. Am. J. Pathol. 160 (6), 2253−2257.

Banner, K.H., Cattaneo, C., Le Net, J.L., Popovic, A., Collins, D., Gale, J.D., 2004. Macroscopic, microscopic and biochemical characterisation of spontaneous colitis in a transgenic mouse, deficient in the multiple drug resistance 1a gene. Br. J. Pharmacol. 143 (5), 590−598.

Barber, J.C., Maloney, V., Hollox, E.J., Stuke-Sontheimer, A., du Bois, G., Daumiller, E., Klein-Vogler, U., Dufke, A., Armour, J.A., Liehr, T., 2005. Duplications and copy number variants of 8p23.1 are cytogenetically indistinguishable but distinct at the molecular level. Eur. J. Hum. Genet. 13 (10), 1131−1136.

Barnett, M., Young, W., Cooney, J., Roy, N., 2014. Metabolomics and proteomics, and what to do with all these 'omes': insights from nutrigenomic investigations in New Zealand. J. Nutr. Nutr. 7 (4−6), 274−282.

Barnett, M.P., Cooney, J.M., Dommels, Y.E., Nones, K., Brewster, D.T., Park, Z., Butts, C.A., McNabb, W.C., Laing, W.A., Roy, N.C., 2013. Modulation of colonic inflammation in Mdr1a$^{(-/-)}$ mice by green tea polyphenols and their effects on the colon transcriptome and proteome. J. Nutr. Biochem. 24 (10), 1678−1690.

Barnett, M.P., McNabb, W.C., Cookson, A.L., Zhu, S., Davy, M., Knoch, B., Nones, K., Hodgkinson, A.J., Roy, N.C., 2010. Changes in colon gene expression associated with increased colon inflammation in interleukin-10 gene-deficient mice inoculated with *Enterococcus* species. BMC Immunol. 11, 39.

Bassett, S.A., Young, W., Barnett, M.P., Cookson, A.L., McNabb, W.C., Roy, N.C., 2015. Changes in composition of caecal microbiota associated with increased colon inflammation in interleukin-10 gene-deficient mice inoculated with *Enterococcus* species. Nutrients 7 (3), 1798−1816.

Baynam, G., Walters, M., Claes, P., Kung, S., LeSouef, P., Dawkins, H., Bellgard, M., Girdea, M., Brudno, M., Robinson, P., Zankl, A., Groza, T., Gillett, D., Goldblatt, J., 2015. Phenotyping: targeting genotype's rich cousin for diagnosis. J. Paediatr. Child Health 51 (4), 381−386.

Bennike, T., Birkelund, S., Stensballe, A., Andersen, V., 2014. Biomarkers in inflammatory bowel diseases: current status and proteomics identification strategies. World J. Gastroenterol 20 (12), 3231–3244.

Berg, D.J., Davidson, N., Kuhn, R., Muller, W., Menon, S., Holland, G., Thompson-Snipes, L., Leach, M.W., Rennick, D., 1996. Enterocolitis and colon cancer in interleukin-10-deficient mice are associated with aberrant cytokine production and CD4(+) TH1-like responses. J. Clin. Invest. 98 (4), 1010–1020.

Beser, O.F., Conde, C.D., Serwas, N.K., Cokugras, F.C., Kutlu, T., Boztug, K., Erkan, T., 2015. Clinical features of interleukin 10 receptor gene mutations in children with very early-onset inflammatory bowel disease. J. Pediatr. Gastroenterol. Nutr. 60 (3), 332–338.

Bilsborough, J., Viney, J.L., 2004. Out, out darn toxin: the role of MDR in intestinal homeostasis. Gastroenterology 127 (1), 339–340.

Bingham, S.A., Gill, C., Welch, A., Day, K., Cassidy, A., Khaw, K.T., Sneyd, M.J., Key, T.J., Roe, L., Day, N.E., 1994. Comparison of dietary assessment methods in nutritional epidemiology: weighed records v. 24 h recalls, food-frequency questionnaires and estimated-diet records. Br. J. Nutr. 72 (4), 619–643.

Boeing, H., 2013. Nutritional epidemiology: new perspectives for understanding the diet-disease relationship? Eur. J. Clin. Nutr. 67 (5), 424–429.

Brahmachary, M., Guilmatre, A., Quilez, J., Hasson, D., Borel, C., Warburton, P., Sharp, A.J., 2014. Digital genotyping of macrosatellites and multicopy genes reveals novel biological functions associated with copy number variation of large tandem repeats. PLoS Genet. 10 (6), e1004418.

Brant, S.R., Panhuysen, C.I., Nicolae, D., Reddy, D.M., Bonen, D.K., Karaliukas, R., Zhang, L., Swanson, E., Datta, L.W., Moran, T., Ravenhill, G., Duerr, R.H., Achkar, J.P., Karban, A.S., Cho, J.H., 2003. MDR1 Ala893 polymorphism is associated with inflammatory bowel disease. Am. J. Hum. Genet. 73 (6), 1282–1292.

Canuel, V., Rance, B., Avillach, P., Degoulet, P., Burgun, A., 2015. Translational research platforms integrating clinical and omics data: a review of publicly available solutions. Brief. Bioinform. 16 (2), 280–290.

Cao, M., Shi, J., Wang, J., Hong, J., Cui, B., Ning, G., 2015. Analysis of human triallelic SNPs by next-generation sequencing. Ann. Hum. Genet. 79 (4), 275–281.

Castro-Santos, P., Suarez, A., Lopez-Rivas, L., Mozo, L., Gutierrez, C., 2006. TNFalpha and IL-10 gene polymorphisms in inflammatory bowel disease. Association of -1082 AA low producer IL-10 genotype with steroid dependency. Am. J. Gastroenterol. 101 (5), 1039–1047.

Chahar, S., Gandhi, V., Yu, S., Desai, K., Cowper-Sal-lari, R., Kim, Y., Perekatt, A.O., Kumar, N., Thackray, J.K., Musolf, A., Kumar, N., Hoffman, A., Londono, D., Vazquez, B.N., Serrano, L., Shin, H., Lupien, M., Gao, N., Verzi, M.P., 2014. Chromatin profiling reveals regulatory network shifts and a protective role for hepatocyte nuclear factor 4alpha during colitis. Mol. Cell Biol. 34 (17), 3291–3304.

Chaisson, M.J., Huddleston, J., Dennis, M.Y., Sudmant, P.H., Malig, M., Hormozdiari, F., Antonacci, F., Surti, U., Sandstrom, R., Boitano, M., Landolin, J.M., Stamatoyannopoulos, J.A., Hunkapiller, M.W., Korlach, J., Eichler, E.E., 2015. Resolving the complexity of the human genome using single-molecule sequencing. Nature 517 (7536), 608–611.

Cheng, C., Pounds, S., 2007. False discovery rate paradigms for statistical analyses of microarray gene expression data. Bioinformation 1 (10), 436–446.

Cohen, S.L., Moore, A.M., Ward, W.E., 2005. Flaxseed oil and inflammation-associated bone abnormalities in interleukin-10 knockout mice. J. Nutr. Biochem. 16 (6), 368–374.

Cooney, J.M., Barnett, M.P., Brewster, D., Knoch, B., McNabb, W.C., Laing, W.A., Roy, N.C., 2012. Proteomic analysis of colon tissue from interleukin-10 gene-deficient mice fed polyunsaturated fatty acids with comparison to transcriptomic analysis. J. Proteome Res. 11 (2), 1065–1077.

Cooney, J.M., Barnett, M.P.G., Dommels, Y.E., Brewster, D., Butts, C.A., McNabb, W.C., Laing, W.A., Roy, N.C., 2016. A combined omics approach to evaluate the effects of dietary curcumin on colon inflammation in the Mdr1a$^{-/-}$ mouse model of inflammatory bowel disease. J. Nutr. Biochem. 27, 181–192.

Coughlin, S.S., 2014. Toward a road map for global -omics: a primer on -omic technologies. Am. J. Epidemiol. 180 (12), 1188–1195.

Dawiskiba, T., Deja, S., Mulak, A., Zabek, A., Jawien, E., Pawelka, D., Banasik, M., Mastalerz-Migas, A., Balcerzak, W., Kaliszewski, K., Skora, J., Barc, P., Korta, K., Pormanczuk, K., Szyber, P., Litarski, A., Mlynarz, P., 2014. Serum and urine metabolomic fingerprinting in diagnostics of inflammatory bowel diseases. World J. Gastroenterol. WJG 20 (1), 163–174.

De Preter, V., Verbeke, K., 2013. Metabolomics as a diagnostic tool in gastroenterology. World J. Gastrointest. Pharmacol. Ther. 4 (4), 97–107.

Delio, M., Patel, K., Maslov, A., Marion, R.W., McDonald, T.V., Cadoff, E.M., Golden, A., Greally, J.M., Vijg, J., Morrow, B., Montagna, C., 2015. Development of a targeted multi-disorder high-throughput sequencing assay for the effective identification of disease-causing variants. PLoS One 10 (7), e0133742.

Delvaux, N., da Costa, V.D., da Costa, M.M., Lampe, E., 2015. Comparison of four methods of genotyping IL28B polymorphisms in chronic hepatitis C patients. J. Virol Methods 220, 1–4.

Denizot, J., Desrichard, A., Agus, A., Uhrhammer, N., Dreux, N., Vouret-Craviari, V., Hofman, P., Darfeuille-Michaud, A., Barnich, N., 2015. Diet-induced hypoxia responsive element demethylation increases CEACAM6 expression, favouring Crohn's disease-associated *Escherichia coli* colonisation. Gut 64 (3), 428–437.

Di Palma, S., Hennrich, M.L., Heck, A.J., Mohammed, S., 2012. Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. J. Proteomics 75 (13), 3791–3813.

Dommels, Y.E.M., Butts, C., Zhu, S.T., Davy, M., Martell, S., Hedderly, D., Barnett, M.P.G., Broadley, K., Roy, N.C., 2007. Characterization of inflammation in mdr1a$^{-/-}$ mice and identification of related gene expression changes in colonic epithelial cells. Genes Nutr. 2 (2), 209–223.

Druart, C., Alligier, M., Salazar, N., Neyrinck, A.M., Delzenne, N.M., 2014. Modulation of the gut microbiota by nutrients with prebiotic and probiotic properties. Adv. Nutr. 5 (5), 624S–633S.

Dryland, P.A., Love, D.R., Walker, M.F., Dommels, Y., Butts, C., Rowan, D.D., Roy, N.C., Helsby, N., Browning, B.L., Zhu, S., Copp, B.R., Ferguson, L.R., 2008. Allantoin as a biomarker of inflammation in an inflammatory bowel disease mouse model: NMR analysis of urine. Open Bioact. Compd. J. 4, 1–6.

Duffus, J.H., Nordberg, M., Templeton, D.M., 2007. IUPAC glossary of terms used in toxicology. Pure Appl. Chem. 79, 192.

Eijssen, L.M., Goelela, V.S., Kelder, T., Adriaens, M.E., Evelo, C.T., Radonjic, M., 2015. A user-friendly workflow for analysis of Illumina gene expression bead array data available at the arrayanalysis.org portal. BMC Genomics 16, 482.

Evelo, C.T., van Bochove, K., Saito, J.T., 2011. Answering biological questions: querying a systems biology database for nutrigenomics. Genes Nutr. 6 (1), 81–87.

Fallaize, R., Forster, H., Macready, A.L., Walsh, M.C., Mathers, J.C., Brennan, L., Gibney, E.R., Gibney, M.J., Lovegrove, J.A., 2014. Online dietary intake estimation: reproducibility and validity of the Food4Me food frequency questionnaire against a 4-day weighed food record. J. Med. Internet Res. 16 (8), e190.

Fan, S., Pan, Z., Geng, Q., Li, X., Wang, Y., An, Y., Xu, Y., Tie, L., Pan, Y., Li, X., 2013. Layered signaling regulatory networks analysis of gene expression involved in malignant tumorigenesis of non-resolving ulcerative colitis via integration of cross-study microarray profiles. PLoS One 8 (6), e67142 [Electronic Resource].

Faye, L.L., Bull, S.B., 2011. Two-stage study designs combining genome-wide association studies, tag single-nucleotide polymorphisms, and exome sequencing: accuracy of genetic effect estimates. BMC Proc. 5 (9), S64.

Ferguson, L.R., 2013. Nutrigenetics, nutrigenomics and inflammatory bowel diseases. Expert Rev. Clin. Immunol. 9 (8), 717−726.

Ferguson, L.R., Laing, B., Marlow, G., Bishop, K.S., 2016. The role of vitamin D in reducing gastrointestinal disease risk and assessment of individual dietary intake needs: focus on genetic and genomic technologies. Mol. Nutr. Food Res. 60 (1), 119−133.

Fiocchi, C., 2014. Integrating omics: the future of IBD? Dig. Dis. 32 (1), 96−102.

Forster, H., Fallaize, R., Gallagher, C., O'Donovan, C.B., Woolhead, C., Walsh, M.C., Macready, A.L., Lovegrove, J.A., Mathers, J.C., Gibney, M.J., Brennan, L., Gibney, E.R., 2014. Online dietary intake estimation: the Food4Me food frequency questionnaire. J. Med. Internet Res. 16 (6), e150.

Franke, A., Balschun, T., Karlsen, T.H., Sventoraityte, J., Nikolaus, S., Mayr, G., Domingues, F.S., Albrecht, M., Nothnagel, M., Ellinghaus, D., Sina, C., Onnie, C.M., Weersma, R.K., Stokkers, P.C., Wijmenga, C., Gazouli, M., Strachan, D., McArdle, W.L., Vermeire, S., Rutgeerts, P., Rosenstiel, P., Krawczak, M., Vatn, M.H., I. s. group, Mathew, C.G., Schreiber, S., 2008. Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. Nat. Genet. 40 (11), 1319−1323.

Franke, A., McGovern, D.P., Barrett, J.C., Wang, K., Radford-Smith, G.L., Ahmad, T., Lees, C.W., Balschun, T., Lee, J., Roberts, R., Anderson, C.A., Bis, J.C., Bumpstead, S., Ellinghaus, D., Festen, E.M., Georges, M., Green, T., Haritunians, T., Jostins, L., Latiano, A., Mathew, C.G., Montgomery, G.W., Prescott, N.J., Raychaudhuri, S., Rotter, J.I., Schumm, P., Sharma, Y., Simms, L.A., Taylor, K.D., Whiteman, D., Wijmenga, C., Baldassano, R.N., Barclay, M., Bayless, T.M., Brand, S., Buning, C., Cohen, A., Colombel, J.F., Cottone, M., Stronati, L., Denson, T., De Vos, M., D'Inca, R., Dubinsky, M., Edwards, C., Florin, T., Franchimont, D., Gearry, R., Glas, J., Van Gossum, A., Guthery, S.L., Halfvarson, J., Verspaget, H.W., Hugot, J.P., Karban, A., Laukens, D., Lawrance, I., Lemann, M., Levine, A., Libioulle, C., Louis, E., Mowat, C., Newman, W., Panes, J., Phillips, A., Proctor, D.D., Regueiro, M., Russell, R., Rutgeerts, P., Sanderson, J., Sans, M., Seibold, F., Steinhart, A.H., Stokkers, P.C., Torkvist, L., Kullak-Ublick, G., Wilson, D., Walters, T., Targan, S.R., Brant, S.R., Rioux, J.D., D'Amato, M., Weersma, R.K., Kugathasan, S., Griffiths, A.M., Mansfield, J.C., Vermeire, S., Duerr, R.H., Silverberg, M.S., Satsangi, J., Schreiber, S., Cho, J.H., Annese, V., Hakonarson, H., Daly, M.J., Parkes, M., 2010. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat. Genet. 42 (12), 1118−1125.

Geiss, G.K., Bumgarner, R.E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D.L., Fell, H.P., Ferree, S., George, R.D., Grogan, T., James, J.J., Maysuria, M., Mitton, J.D., Oliveri, P., Osborn, J.L., Peng, T., Ratcliffe, A.L., Webster, P.J., Davidson, E.H., Hood, L., Dimitrov, K., 2008. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat. Biotechnol. 26 (3), 317−325.

Golzio, C., Katsanis, N., 2013. Genetic architecture of reciprocal CNVs. Curr. Opin. Genet. Dev. 23 (3), 240−248.

Gonzalez, I., Cao, K.A., Davis, M.J., Dejean, S., 2012. Visualising associations between paired 'omics' data sets. BioData Min. 5 (1), 19.

Goyette, P., Boucher, G., Mallon, D., Ellinghaus, E., Jostins, L., Huang, H., Ripke, S., Gusareva, E.S., Annese, V., Hauser, S.L., Oksenberg, J.R., Thomsen, I., Leslie, S., . International Inflammatory Bowel Disease Genetics, C., Australia, I. New Zealand, I. B. D. G. C., Belgium, I. B. D. G. C., Italian Group for, N. I. B. D. G. Consortium, I. United Kingdom, C. Wellcome Trust Case Control, I. B. D. G. C. Quebec, Daly, M.J., Van Steen, K., Duerr, R.H., Barrett, J.C., McGovern, D.P., Schumm, L.P., Traherne, J.A., Carrington, M.N., Kosmoliaptsis, V., Karlsen, T.H., Franke, A., Rioux, J.D., 2015. High-density mapping of the MHC identifies a shared role for HLA-DRB1 01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. Nat. Genet. 47 (2), 172−179.

Greco, L., Romino, R., Coto, I., Di Cosmo, N., Percopo, S., Maglio, M., Paparo, F., Gasperi, V., Limongelli, M.G., Cotichini, R., D'Agate, C., Tinto, N., Sacchetti, L., Tosi, R., Stazi, M.A., 2002. The first large population based twin study of coeliac disease. Gut 50 (5), 624−628.

Hallmann, J., Kolossa, S., Gedrich, K., Celis-Morales, C., Forster, H., O'Donovan, C.B., Woolhead, C., Macready, A.L., Fallaize, R., Marsaux, C.F., Lambrinou, C.P., Mavrogianni, C., Moschonis, G., Navas-Carretero, S., San-Cristobal, R., Godlewska, M., Surwillo, A., Mathers, J.C., Gibney, E.R., Brennan, L., Walsh, M.C., Lovegrove, J.A., Saris, W.H., Manios, Y., Martinez, J.A., Traczyk, I., Gibney, M.J., Daniel, H., S. Food4Me, 2015. Predicting fatty acid profiles in blood based on food intake and the FADS1 rs174546 SNP. Mol. Nutr. Food Res. 59 (12), 2565−2573.

Haraksingh, R.R., Snyder, M.P., 2013. Impacts of variation in the human genome on gene regulation. J. Mol. Biol. 425 (21), 3970−3977.

Haridan, U.S., Mokhtar, U., Machado, L.R., Abdul Aziz, A.T., Shueb, R.H., Zaid, M., Sim, B., Mustafa, M., Nik Yusof, N.K., Lee, C.K., Abu Bakar, S., AbuBakar, S., Hollox, E.J., Boon Peng, H., 2015. A comparison of assays for accurate copy number measurement of the low-affinity Fc gamma receptor genes FCGR3A and FCGR3B. PLoS One 10 (1), e0116791 [Electronic Resource].

Heckman-Stoddard, B.M., Smith, J.J., 2014. Precision medicine clinical trials: defining new treatment strategies. Semin Oncol. Nurs. 30 (2), 109−116.

Hegazi, R.A., Saad, R.S., Mady, H., Matarese, L.E., O'Keefe, S., Kandil, H.M., 2006. Dietary fatty acids modulate chronic colitis, colitis-associated colon neoplasia and COX-2 expression in IL-10 knockout mice. Nutrition 22 (3), 275−282.

Hehir-Kwa, J.Y., Pfundt, R., Veltman, J.A., de Leeuw, N., 2013. Pathogenic or not? Assessing the clinical relevance of copy number variants. Clin. Genet. 84 (5), 415−421.

Ho, G.T., Moodie, F.M., Satsangi, J., 2003. Multidrug resistance 1 gene (P-glycoprotein 170): an important determinant in gastrointestinal disease? Gut 52 (5), 759−766.

Ho, G.T., Soranzo, N., Nimmo, E.R., Tenesa, A., Goldstein, D.B., Satsangi, J., 2006. ABCB1/MDR1 gene determines susceptibility and phenotype in ulcerative colitis: discrimination of critical variants using a gene-wide haplotype tagging approach. Hum. Mol. Genet. 15 (5), 797−805.

Hooper, L., Thompson, R.L., Harrison, R.A., Summerbell, C.D., Ness, A.R., Moore, H.J., Worthington, H.V., Durrington, P.N., Higgins, J.P., Capps, N.E., Riemersma, R.A., Ebrahim, S.B., Davey Smith, G., 2006. Risks and benefits of omega 3 fats for mortality, cardiovascular disease, and cancer: systematic review. BMJ 332 (7544), 752−760.

Huebner, C., Petermann, I., Browning, B.L., Shelling, A.N., Ferguson, L.R., 2007. Triallelic single nucleotide polymorphisms and genotyping error in genetic epidemiology studies: MDR1 (ABCB1) G2677/T/A as an example. Cancer Epidemiol. Biomarkers Prev. 16 (6), 1185−1192.

Ismail, N.A., Posma, J.M., Frost, G., Holmes, E., Garcia-Perez, I., 2013. The role of metabonomics as a tool for augmenting nutritional information in epidemiological studies. Electrophoresis 34 (19), 2776−2786.

Janes, J., Hu, F., Lewin, A., Turro, E., 2015. A comparative study of RNA-seq analysis strategies. Brief. Bioinform. 16 (6), 932−940.

Jin, S.C., Benitez, B.A., Deming, Y., Cruchaga, C., 2016. Pooled-DNA sequencing for elucidating new genomic risk factors, rare variants underlying Alzheimer's disease. Methods Mol. Biol. 1303, 299−314.

Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A., Essers, J., Mitrovic, M., Ning, K., Cleynen, I., Theatre, E., Spain, S.L., Raychaudhuri, S., Goyette, P., Wei, Z., Abraham, C., Achkar, J.P., Ahmad, T., Amininejad, L., Ananthakrishnan, A.N., Andersen, V., Andrews, J.M., Baidoo, L., Balschun, T., Bampton, P.A., Bitton, A., Boucher, G., Brand, S., Buning, C., Cohain, A., Cichon, S., D'Amato, M., De Jong, D., Devaney, K.L., Dubinsky, M., Edwards, C., Ellinghaus, D., Ferguson, L.R., Franchimont, D., Fransen, K., Gearry, R., Georges, M., Gieger, C., Glas, J., Haritunians, T., Hart, A., Hawkey, C., Hedl, M., Hu, X., Karlsen, T.H., Kupcinskas, L., Kugathasan, S., Latiano, A., Laukens, D., Lawrance, I.C., Lees, C.W., Louis, E., Mahy, G., Mansfield, J., Morgan, A.R., Mowat, C., Newman, W., Palmieri, O., Ponsioen, C.Y., Potocnik, U., Prescott, N.J., Regueiro, M., Rotter, J.I., Russell, R.K., Sanderson, J.D., Sans, M., Satsangi, J., Schreiber, S., Simms, L.A., Sventoraityte, J., Targan, S.R., Taylor, K.D., Tremelling, M., Verspaget, H.W., De Vos, M., Wijmenga, C., Wilson, D.C., Winkelmann, J., Xavier, R.J., Zeissig, S., Zhang, B., Zhang, C.K., Zhao, H., I. B. D. G. C. International, Silverberg, M.S., Annese, V., Hakonarson, H., Brant, S.R., Radford-Smith, G., Mathew, C.G., Rioux, J.D., Schadt, E.E., Daly, M.J., Franke, A., Parkes, M., Vermeire, S., Barrett, J.C., Cho, J.H., 2012. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 491 (7422), 119−124.

Joven, J., Micol, V., Segura-Carretero, A., Alonso-Villaverde, C., Menendez, J.A., Bioactive Food Components, P., 2014. Polyphenols and the modulation of gene expression pathways: can we eat our way out of the danger of chronic disease? Crit. Rev. Food Sci. Nutr. 54 (8), 985−1001.

Kang, J., Zhu, L., Lu, J., Zhang, X., 2015. Application of metabolomics in autoimmune diseases: insight into biomarkers and pathology. J. Neuroimmunol. 279, 25−32.

Kaput, J., Dawson, K., 2007. Complexity of type 2 diabetes mellitus data sets emerging from nutrigenomic research: a case for dimensionality reduction? Mutat. Res. 622 (1−2), 19−32.

Kellermayer, R., Balasa, A., Zhang, W., Lee, S., Mirza, S., Chakravarty, A., Szigeti, R., Laritsky, E., Tatevian, N., Smith, C.W., Shen, L., Waterland, R.A., 2010. Epigenetic maturation in colonic mucosa continues beyond infancy in mice. Hum. Mol. Genet. 19 (11), 2168−2176.

Klein, W., Tromm, A., Griga, T., Fricke, H., Folwaczny, C., Hocke, M., Eitner, K., Marx, M., Runte, M., Epplen, J.T., 2000. The IL-10 gene is not involved in the predisposition to inflammatory bowel disease. Electrophoresis 21 (17), 3578−3582.

Knoch, B., Barnett, M.P., Cooney, J., McNabb, W.C., Barraclough, D., Laing, W., Zhu, S., Park, Z.A., Maclean, P., Knowles, S.O., Roy, N.C., 2010a. Molecular characterization of the onset and progression of colitis in inoculated Interleukin-10 gene-deficient mice: a role for PPARalpha. PPAR Res. 2010, 621069.

Knoch, B., Barnett, M.P., McNabb, W.C., Zhu, S., Park, Z.A., Khan, A., Roy, N.C., 2010b. Dietary arachidonic acid-mediated effects on colon inflammation using transcriptome analysis. Mol. Nutr. Food Res. 54 (1), S62−S74.

Knoch, B., Barnett, M.P., Zhu, S., Park, Z.A., Nones, K., Dommels, Y.E., Knowles, S.O., McNabb, W.C., Roy, N.C., 2009. Genome-wide analysis of dietary eicosapentaenoic acid- and oleic acid-induced modulation of colon inflammation in interleukin-10 gene-deficient mice. J. Nutr. Nutr. 2 (1), 9−28.

Konstantinidou, V., Ruiz, L.A., Ordovas, J.M., 2014. Personalized nutrition and cardiovascular disease prevention: from Framingham to PREDIMED. Adv. Nutr. 5 (3), 368S−371S.

Kuhn, R., Lohler, J., Rennick, D., Rajewsky, K., Muller, W., 1993. Interleukin-10-deficient mice develop chronic enterocolitis. Cell 75 (2), 263−274.

Kutmon, M., Coort, S.L., de Nooijer, K., Lemmens, C., Evelo, C.T., 2015. Integrative network-based analysis of mRNA and microRNA expression in 1,25-dihydroxyvitamin D3-treated cancer cells. Genes Nutr. 10 (5), 484.

Lappalainen, M., Halme, L., Turunen, U., Saavalainen, P., Einarsdottir, E., Farkkila, M., Kontula, K., Paavola-Sakki, P., 2008. Association of IL23R, TNFRSF1A, and HLA-DRB1*0103 allele variants with inflammatory bowel disease phenotypes in the Finnish population. Inflamm. Bowel Dis. 14 (8), 1118−1124.

Lin, H.M., Barnett, M.P., Roy, N.C., Joyce, N.I., Zhu, S., Armstrong, K., Helsby, N.A., Ferguson, L.R., Rowan, D.D., 2010. Metabolomic analysis identifies inflammatory and noninflammatory metabolic effects of genetic modification in a mouse model of Crohn's disease. J. Proteome Res. 9 (4), 1965−1975.

Lin, H.M., Edmunds, S.J., Zhu, S., Helsby, N.A., Ferguson, L.R., Rowan, D.D., 2011. Metabolomic analysis reveals differences in urinary excretion of kiwifruit-derived metabolites in a mouse model of inflammatory bowel disease. Mol. Nutr. Food Res. 55 (12), 1900−1904.

Liu, Y., Wei, X., Kong, X., Guo, X., Sun, Y., Man, J., Du, L., Zhu, H., Qu, Z., Tian, P., Mao, B., Yang, Y., 2015. Targeted next-generation sequencing for clinical diagnosis of 561 mendelian diseases. PLoS One 10 (8), e0133636.

Livingstone, K.M., Celis-Morales, C., Lara, J., Ashor, A.W., Lovegrove, J.A., Martinez, J.A., Saris, W.H., Gibney, M., Manios, Y., Traczyk, I., Drevon, C.A., Daniel, H., Gibney, E.R., Brennan, L., Bouwman, J., Grimaldi, K.A., Mathers, J.C., 2015. Associations between FTO genotype and total energy and macronutrient intake in adults: a systematic review and meta-analysis. Obes. Rev. 16 (8), 666−678.

Love, M., Anders, S., Kim, V., Huber, W., 2015. RNA-seq Workflow: Gene-level Exploratory Analysis and Differential Expression. http://www.bioconductor.org/help/workflows/rnaseqGene/.

Lu, Y., Ek, W.E., Whiteman, D., Vaughan, T.L., Spurdle, A.B., Easton, D.F., Pharoah, P.D., Thompson, D.J., Dunning, A.M., Hayward, N.K., Chenevix-Trench, G., Q-MEGA and AMFS Investigators, Investigators, A., ANECS, S., UKOPS, S., Consortium, B., Macgregor, S., 2014. Most common 'sporadic' cancers have a significant germline genetic component. Hum. Mol. Genet. 23 (22), 6112−6118.

MacGregor, A.J., Snieder, H., Schork, N.J., Spector, T.D., 2000. Twins. Novel uses to study complex traits and genetic diseases. Trends Genet. 16 (3), 131−134.

Marchesi, J.R., Adams, D.H., Fava, F., Hermes, G.D., Hirschfield, G.M., Hold, G., Quraishi, M.N., Kinross, J., Smidt, H., Tuohy, K.M., Thomas, L.V., Zoetendal, E.G., Hart, A., 2016. The gut microbiota and host health: a new clinical frontier. Gut 65 (2), 330−339.

Marlow, G., Ellett, S., Ferguson, I.R., Zhu, S., Karunasinghe, N., Jesuthasan, A.C., Han, D.Y., Fraser, A.G., Ferguson, L.R., 2013. Transcriptomics to study the effect of a Mediterranean-inspired diet on inflammation in Crohn's disease patients. Hum. Genomics 7, 24.

Martin, H., Burgess, E.J., Smith, W.A., McGhie, T.K., Cooney, J.M., Lunken, R.C., de Guzman, E., Trower, T., Perry, N.B., 2015. JAK2 and AMP-kinase inhibition in vitro by food extracts, fractions and purified phytochemicals. Food Funct. 6 (1), 305−312.

Mazaika, E., Homsy, J., 2014. Digital droplet PCR: CNV analysis and other applications. Curr. Protoc. Hum. Genet. 82, 7.24.21−27.24.13.

McCarthy, J., O'Mahony, L., O'Callaghan, L., Sheil, B., Vaughan, E.E., Fitzsimons, N., Fitzgibbon, J., O'Sullivan, G.C., Kiely, B., Collins, J.K., Shanahan, F., 2003. Double blind, placebo controlled trial of two probiotic strains in interleukin 10 knockout mice and mechanistic link with cytokine balance. Gut 52 (7), 975−980.

Metzker, M.L., 2010. Sequencing technologies − the next generation. Nat. Rev. Genet. 11 (1), 31−46.

Moller, P., Loft, S., 2004. Interventions with antioxidants and nutrients in relation to oxidative DNA damage and repair. Mutat. Res. 551 (1−2), 79−89.

Moller, P., Loft, S., 2006. Dietary antioxidants and beneficial effect on oxidatively damaged DNA. Free Radic. Biol. Med. 41 (3), 388−415.

Nanau, R.M., Neuman, M.G., 2012. Metabolome and inflammasome in inflammatory bowel disease. Transl. Res. J. Lab. Clin. Med. 160 (1), 1−28.

Nasef, N.A., Mehta, S., Murray, P., Marlow, G., Ferguson, L.R., 2014. Anti-inflammatory activity of fruit fractions in vitro, mediated through toll-like receptor 4 and 2 in the context of inflammatory bowel disease. Nutrients 6 (11), 5265−5279.

Neurath, M.F., 2012. Animal models of inflammatory bowel diseases: illuminating the pathogenesis of colitis, ileitis and cancer. Dig. Dis. 30 (1), 91−94.

Nones, K., Dommels, Y.E., Martell, S., Butts, C., McNabb, W.C., Park, Z.A., Zhu, S., Hedderley, D., Barnett, M.P., Roy, N.C., 2009. The effects of dietary curcumin and rutin on colonic inflammation and gene expression in multidrug resistance gene-deficient (mdr1a$^{-/-}$) mice, a model of inflammatory bowel diseases. Br. J. Nutr. 101 (2), 169−181.

Panwala, C.M., Jones, J.C., Viney, J.L., 1998. A novel model of inflammatory bowel disease: mice deficient for the multiple drug resistance gene, mdr1a, spontaneously develop colitis. J. Immunol. 161 (10), 5733−5744.

Park, J.M., Han, N.Y., Han, Y.M., Chung, M.K., Lee, H.K., Ko, K.H., Kim, E.H., Hahm, K.B., 2014. Predictive proteomic biomarkers for inflammatory bowel disease-associated cancer: where are we now in the era of the next generation proteomics? World J. Gastroenterol. 20 (37), 13466−13476.

Paturi, G., Mandimika, T., Butts, C.A., Zhu, S., Roy, N.C., McNabb, W.C., Ansell, J., 2012. Influence of dietary blueberry and broccoli on cecal microbiota activity and colon morphology in mdr1a$^{(-/-)}$ mice, a model of inflammatory bowel diseases. Nutrition 28 (3), 324−330.

Peterson, C.T., Sharma, V., Elmen, L., Peterson, S.N., 2015. Immune homeostasis, dysbiosis and therapeutic modulation of the gut microbiota. Clin. Exp. Immunol. 179 (3), 363−377.

Polytarchou, C., Koukos, G., Iliopoulos, D., 2014. Systems biology in inflammatory bowel diseases: ready for prime time. Curr. Opin. Gastroenterol. 30 (4), 339−346.

Prattis, S., Jurjus, A., 2015. Spontaneous and transgenic rodent models of inflammatory bowel disease. Lab. Anim. Res. 31 (2), 47−68.

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., Gonzalez, J.R., Gratacos, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W., Hurles, M.E., 2006. Global variation in copy number in the human genome. Nature 444 (7118), 444−454.

Roberts, R.L., Gearry, R.B., Hollis-Moffatt, J.E., Miller, A.L., Reid, J., Abkevich, V., Timms, K.M., Gutin, A., Lanchbury, J.S., Merriman, T.R., Barclay, M.L., Kennedy, M.A., 2007. IL23R R381Q and ATG16L1 T300A are strongly associated with Crohn's disease in a study of New Zealand Caucasians with inflammatory bowel disease. Am. J. Gastroenterol. 102 (12), 2754−2761.

Roy, N.C., Barnett, M.P.G., Knoch, B., Dommels, Y.E.M., McNabb, W.C., 2007. Nutrigenomics applied to an animal model of Inflammatory Bowel Diseases: transcriptomic analysis of the effects of eicosapentaenoic acid- and arachidonic acid-enriched diets. Mutat. Res. 622 (1−2), 103−116.

Russ, A.E., Peters, J.S., McNabb, W.C., Barnett, M.P., Anderson, R.C., Park, Z., Zhu, S., Maclean, P., Young, W., Reynolds, G.W., Roy, N.C., 2013. Gene expression changes in the colon epithelium are similar to those of intact colon during late inflammation in interleukin-10 gene deficient mice. PLoS One 8 (5), e63251.

Serban, C., Dragan, S., 2014. The relationship between inflammatory and oxidative stress biomarkers, atherosclerosis and rheumatic diseases. Curr. Pharm. Des. 20 (4), 585−600.

Shen, F., Huang, J., Fitch, K.R., Truong, V.B., Kirby, A., Chen, W., Zhang, J., Liu, G., McCarroll, S.A., Jones, K.W., Shapero, M.H., 2008. Improved detection of global copy number variation using high density, non-polymorphic oligonucleotide probes. BMC Genet. 9, 27.

Shen, W., Gaskins, H.R., McIntosh, M.K., 2014. Influence of dietary fat on intestinal microbes, inflammation, barrier function and metabolic outcomes. J. Nutr. Biochem. 25 (3), 270−280.

Shi, C.Z., Chen, H.Q., Liang, Y., Xia, Y., Yang, Y.Z., Yang, J., Zhang, J.D., Wang, S.H., Liu, J., Qin, H.L., 2014. Combined probiotic bacteria promotes intestinal epithelial barrier function in interleukin-10-gene-deficient mice. World J. Gastroenterol. 20 (16), 4636−4647.

Smyth, G.K., 2005. Limma: linear models for microarray data. In: Gentleman, R., Carey, V., Dudoit, S., Irizarry, R., Huber, W. (Eds.), Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Springer, New York, pp. 397−420.

Stroeve, J.H., van Wietmarschen, H., Kremer, B.H., van Ommen, B., Wopereis, S., 2015. Phenotypic flexibility as a measure of health: the optimal nutritional stress response test. Genes Nutr. 10 (3), 459.

Triggs, C.M., Munday, K., Hu, R., Fraser, A.G., Gearry, R.B., Barclay, M.L., Ferguson, L.R., 2010. Dietary factors in chronic inflammation: food tolerances and intolerances of a New Zealand Caucasian Crohn's disease population. Mutat. Res. 690 (1−2), 123−138.

Ulluwishewa, D., Anderson, R.C., Young, W., McNabb, W.C., van Baarlen, P., Moughan, P.J., Wells, J.M., Roy, N.C., 2015. Live *Faecalibacterium prausnitzii* in an apical anaerobic model of the intestinal epithelial barrier. Cell Microbiol. 17 (2), 226−240.

Valatas, V., Vakas, M., Kolios, G., 2013. The value of experimental models of colitis in predicting efficacy of biological therapies for inflammatory bowel diseases. Am. J. Physiol. Gastrointest. Liver Physiol. 305 (11), G763−G785.

van Iersel, M.P., Kelder, T., Pico, A.R., Hanspers, K., Coort, S., Conklin, B.R., Evelo, C., 2008. Presenting and exploring biological pathways with PathVisio. BMC Bioinf. 9, 399.

VanderWeele, T.J., Tchetgen Tchetgen, E.J., Cornelis, M., Kraft, P., 2014. Methodological challenges in mendelian randomization. Epidemiology 25 (3), 427−435.

Wang, G.T., Zhang, D., Li, B., Dai, H., Leal, S.M., 2015a. Collapsed haplotype pattern method for linkage analysis of next-generation sequence data. Eur. J. Hum. Genet. 23 (12), 1739−1743.

Wang, Q., Lu, Q., Zhao, H., 2015b. A review of study designs and statistical methods for genomic epidemiology studies using next generation sequencing. Front. Genet. 6, 149.

Wang, H., Xue, Y., Zhang, H., Huang, Y., Yang, G., Du, M., Zhu, M.J., 2013. Dietary grape seed extract ameliorates symptoms of inflammatory bowel disease in IL10-deficient mice. Mol. Nutr. Food Res. 57 (12), 2253−2257.

Weersma, R.K., Zhernakova, A., Nolte, I.M., Lefebvre, C., Rioux, J.D., Mulder, F., van Dullemen, H.M., Kleibeuker, J.H., Wijmenga, C., Dijkstra, G., 2008. ATG16L1 and IL23R are associated with inflammatory bowel diseases but not with celiac disease in the Netherlands. Am. J. Gastroenterol. 103 (3), 621−627.

Wellcome Trust Case Control, C, 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447 (7145), 661−678.

Wilk, J.N., Bilsborough, J., Viney, J.L., 2005. The mdr1a$^{-/-}$ mouse model of spontaneous colitis: a relevant and appropriate animal model to study inflammatory bowel disease. Immunol. Res. 31 (2), 151−159.

Willett, W.C., 1995. Dietary assessment methods. Br. J. Nutr. 74 (1), 141−143.

Wu, Q., Yuan, H., Zhang, L., Zhang, Y., 2012. Recent advances on multidimensional liquid chromatography-mass spectrometry for proteomics: from qualitative to quantitative analysis−a review. Anal. Chim. Acta 731, 1−10.

Xu, C., Aragam, N., Li, X., Villla, E.C., Wang, L., Briones, D., Petty, L., Posada, Y., Arana, T.B., Cruz, G., Mao, C., Camarillo, C., Su, B.B., Escamilla, M.A., Wang, K., 2013. BCL9 and C9orf5 are associated with negative symptoms in schizophrenia: meta-analysis of two genome-wide association studies. PLoS One 8 (1), e51674 [Electronic Resource].

Yang, G., Xue, Y., Zhang, H., Du, M., Zhu, M.J., 2015. Favourable effects of grape seed extract on intestinal epithelial differentiation and barrier function in IL10-deficient mice. Br. J. Nutr. 114 (1), 15−23.

Yen, D., Cheung, J., Scheerens, H., Poulet, F., McClanahan, T., McKenzie, B., Kleinschek, M.A., Owyang, A., Mattson, J., Blumenschein, W., Murphy, E., Sathe, M., Cua, D.J., Kastelein, R.A., Rennick, D., 2006. IL-23 is essential for T cell-mediated colitis and promotes inflammation via IL-17 and IL-6. J. Clin. Invest. 116 (5), 1310−1316.

Yoder, S.J., 2015. ChIP-on-chip analysis methods for affymetrix tiling arrays. Methods Mol. Biol. 1288, 473−487.

Zhang, J.X., Song, J., Wang, J., Dong, W.G., 2014. JAK2 rs10758669 polymorphisms and susceptibility to ulcerative colitis and Crohn's disease: a meta-analysis. Inflammation 37 (3), 793−800.

Zhao, S., Fung-Leung, W.P., Bittner, A., Ngo, K., Liu, X., 2014. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One 9 (1), e78644.

Chapter 18

DNA Microarrays and Genetic Testing

A. Deshpande and P.S. White
Los Alamos National Laboratory, Los Alamos, NM, United States

18.1 INTRODUCTION

Since DNA microarrays were first introduced in the early 1990s, applications of this technology have expanded beyond understanding gene expression anomalies in disease and identifying biomarkers to routine use in diagnosis and prognosis. Moreover, as this is a detection platform, innovations in types of arrays have increased the diversity of markers that can be probed for, leading to applications in genetic testing that range from cancer staging and risk estimation of cancer and Mendelian diseases to diagnosis of autism spectrum disorders and prenatal disease diagnosis. The ever-expanding information on the human genome, transcriptome, and metabolome has further facilitated the adoption of this technology to custom biomarker detection for various applications.

DNA microarrays are platforms that allow for the simultaneous detection of multiple nucleic acid-based signatures (sequences unique to the target organism being searched for) in a single sample. Traditional microarrays have been two-dimensional solid surfaces (glass, plastic, or silicon) that have multiple (hundreds of thousands to millions) of DNA fragments/genomic DNA covalently linked to them. Upon exposure to a sample of interest, these probes hybridize to a target that has been labeled with a fluorophore to allow detection. A classical microarray application involves four major steps: sample labeling (following nucleic acid extraction), hybridization to the microarray platform, signal detection by imaging, and data analysis and interpretation. This technology has revolutionized how we study biological processes and identify and apply biomarkers for disease diagnosis and prognosis. DNA microarray technology is now mature and routinely used with several commercial manufacturers of standard and custom microarrays (Agilent, Affymetrix, Illumina, etc., Stanford microarray database). A new type of DNA microarray platform has also emerged: the three-dimensional array, which uses micro- and even nanometer-sized microspheres that allow for the covalent linkage of probes and detection with flow cytometry, microscopy, or other detection instruments (Rödiger et al., 2014). These are commonly called suspension arrays and offer different features that solid microarrays do not. A cursory literature search for the keywords "microsphere arrays AND genetic testing" between 2010–15 yields a total of 17,500 publications in just 5 years, versus 17,200 publications for the keywords "microarrays AND genetic testing" for the same time frame. Given the relatively new arrival of microsphere arrays, it seems that suspension arrays have gained broad acceptance and application in a shorter period of time than solid phase microarrays that have been in existence for approximately three decades.

The key feature of microarrays is that they are a type of detection platform that facilitates multiplexing or the simultaneous detection of multiple markers in a single sample. A diagnostic assay has the following broad components: one or more signatures that the assay probes for, an assay chemistry that finds that signature using one or more probes, a detection format that allows for the reporting of the found signature, and data analysis that takes the reported signal from the detection format, processes it, and provides a definitive result for the sample. Multiplexing can be accomplished at the assay chemistry level or the detection platform level. Microarrays can serve as both the platform where assay chemistry, such as hybridization, detects a signature and a platform that captures the product of a separate signature detecting multiplexed assay chemistry, such as a polymerase chain reaction (PCR) or ligation. This versatility of microarrays lends itself to diverse applications as well as diverse biomarkers. The

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00018-3
Copyright © 2017 Elsevier Ltd. All rights reserved.

field of cancer research contributed immensely to the identification of gene expression biomarkers that can be used to study cancer progression or diagnose cancer (Kulasingam and Diamandis, 2008; Trevino et al., 2007). Today, microarrays are no longer merely used for gene expression analysis but also for the detection of diverse genetic alterations such as insertions, deletions, single nucleotide variants (SNPs), repeats, alternative splice junctions, and alterations in DNA methylation patterns (Trevino et al., 2007).

This chapter will cover the different types of solid and suspension DNA arrays and their application to genetic testing in humans. It is meant to provide a broad overview rather than focusing on one specific area, such as cancer, as the assay chemistries, types of signatures, and detection chemistries can be generalized across different applications. Moreover, due to the maturity of this field, applications of microarrays have become diverse. It seems relevant to provide the reader with a glimpse of the breadth currently in existence along with examples of new innovations in the types of assays that work toward making this technology both high throughput and multiplex capable at lower costs. While today's microarrays can use both RNA and DNA as probes for detection, this overview will describe applications that use DNA probes.

18.2 DNA MICROARRAYS

This section will cover information on both solid support and suspension arrays and concludes with a comparison between the two broad types. The data analysis component of these technologies will not be discussed in detail, but readers can refer to Allison et al., (2006) or Drăghici (2003) for information on this topic. Both detection platforms have their own methods of normalization and data processing, and due to the mature nature of the field, several companies provide data analysis packages that go with the detection instrument. Fundamentally, both platforms primarily use fluorescence detection, some form of normalization using background signals, and either ratiometric or absolute value reporting. Several statistical packages are further used for data analysis, depending upon the application of the platform to diagnostics or understanding the biology of disease.

18.2.1 Solid Phase or Two-Dimensional DNA Microarrays

The roots of DNA microarrays can be traced back to techniques from the 1970s of replica plating and generating DNA spots on nitrocellulose paper, cloned DNA fragments that were detected using radiolabeled probes (Bumgarner, 2013). In this version, the sample of interest was immobilized, and the detection probe was in suspension, but the assay chemistry was still hybridization. DNA microarrays of today screen a sample by immobilizing the detection probe of a known sequence, and the unknown target is in suspension. Thus the concept of screening a sample with multiple probes simultaneously was common to both techniques. The evolution of better detection techniques using fluorescence, better methods of manufacturing solid platforms for DNA probes, and the explosion of genetic sequence data together led to the development of the modern day two-dimensional microarrays that are routinely used in clinical practice. Bumgarner (2013) provides an excellent, concise description of the history of DNA microarrays.

Solid phase or two-dimensional microarrays can be broadly divided into three types, depending on their function: quantification of gene expression, detection of a specific genetic signature like an SNP or a unique sequence, or a change in the quantity of a signature like copy number variation or aneuploidy. Fig. 18.1 shows an example of the typical steps found in two-dimensional microarray analysis that include nucleic acid extraction, fluorescent labeling, hybridization to the array, and readout by scanning for fluorescence. The underlying chemistry is hybridization, and the type of probe immobilized on the solid support determines what the microarray can be used for. Microarrays can be functionalized with oligonucleotides (cloned or synthesized), cDNA, or PCR products by either spotting on glass slides or direct synthesis on the chip, as is done by some commercial companies.

Quantification of gene expression can be performed by directly measuring the hybridized target to a probe (one sample system; Dyrskjøt et al., 2010) or through a comparison between the expression of the test sample and a reference sample. This is called comparative hybridization. Both methods employ hybridization for their assay chemistry. Arrays are functionalized (covalently linked) with cRNA-specific oligonucleotides, and the total RNA is extracted from test samples (and reference samples) and reverse transcribed to produce labeled cDNA or subsequently transcribed to produce labeled cRNA. Probes can be either cloned (bacterial artificial chromosomes) or synthetic oligonucleotides or PCR products. The product (cDNA or cRNA) is hybridized to the array, following which the array is scanned and fluorescent spots identified. In the case of comparative hybridization, the labeling of the test and reference sample cDNA or cRNA can be done using different colored fluorescent dyes, and a fluorescence ratio is calculated to determine expression differences between a reference and test sample. This method looks for upregulation or downregulation of gene expression and can be used to study the biology of disease progression like cancer or the pathogenesis of hypertrophic cardiomyopathy (Bos et al., 2009). This assay does not require any amplification, and this type of a chemistry is useful for understanding the biology of a disease or discovering biomarkers that can be used for genetic disease diagnostics,

FIGURE 18.1 An example of steps that occur during the process of DNA microarray analysis: nucleic acid extraction, labeling, hybridization, and scanning to read the fluorescence (https://en.wikipedia.org/wiki/DNA_microarray#/media/File:Microarray-schema.jpg public domain, https://en.wikipedia.org/w/index.php?curid=13359806https://en.wikipedia.org/wiki/File:Microarray-schema.jpg#file).

understanding the etiology of cancer, staging cancers, and monitoring disease progression.

Changes in the copy number of a gene or chromosome number can also be detected by the same comparative method commonly called comparative genomic hybridization (CGH), and the microarrays are commonly known as chromosomal microarrays (CMAs). Array CGH is routinely used to compare the copy number variation in cancerous tissue with normal tissue when DNA from both types is run on microarrays. CMAs have seen application primarily in cancer and autism spectrum disorder detection (Wapner et al., 2012). In fact, CMAs have become so mature these days that according to some, they should be considered as part of the initial diagnostic evaluation of patients with autism spectrum disorders (Shen et al., 2010). Miller et al. (2010) conducted a comparative analysis between CMA and G-banded karyotyping across 33 different studies to understand the limitations and benefits of using CMA over standard cytogenetic techniques to diagnose individuals with developmental disabilities and congenital anomalies. They concluded that CMA has a significantly higher diagnostic yield and recommended that this technique be used as a first-tier cytogenetic test and replace G-banded karyotyping for patients that do not

show obvious signs of a recognizable syndrome. They developed an algorithm and decision tree that can be applied following the use of CMAs for patients with unexplained developmental disabilities or congenital abnormalities.

SNP arrays have been used for determining disease susceptibility and for measuring the efficacy of customized drug therapies and even loss of heterozygosity analysis in cancer. In addition, SNP arrays are now being used to detect copy number variation (Mao et al., 2007; Framboise, 2009). Probes consist of allele-specific sequences with the SNP of interest included in the probe. An amplification step of the DNA in a test sample such as multiplex PCR is almost always used to enhance the sensitivity of SNP detection by the microarray. Advances in amplification include whole genome amplification following a restriction digest of the genomic DNA (Kennedy et al., 2003; Matsuzaki et al., 2004). The amplified product that is also labeled is then allowed to hybridize to various allele-specific probes on the array. Alternatively, a single base extension (SBE) is performed via PCR amplification of specific SNP-containing regions, fragmentation of the product and array-based hybridization, and enzymatic SBE on the array using fluorescently labeled dideoxy nucleotide phosphates (ddNTPs) and signal detection (Kurg et al., 2000; Steemers et al., 2006). SBE can also be conducted in solution prior to capture on the microarrays (Milani et al., 2006). SBE allows for the application of SNP arrays to DNA samples that may be degraded and not amenable to restriction digestion like those from paraffin-embedded tissue samples. The method described above requires the design of multiple PCR primers for all targets of interest as well as separate hybridization probes for the array for SBE. In another clever variation, a 2-fold PCR reaction (multiplex followed by universal) is conducted using the same primer pair designed to have modular components that first bind to the target genomic DNA sequences immediately upstream of the SNP site. This reaction generates multiple copies of the SNP-containing target and a universal primer pair complement. Following universal PCR that amplifies all the SNP targets using a single primer pair, the products are hybridized using the first primer pair oligos now hybridized to arrays, and SBE is conducted using four colored ddNTPs and detected (Krjutskov et al., 2008). This approach has two main advantages over the standard SBE process. First, as the PCR products hybridized to the microarray are shorter, there is no need for a fragmentation step. Second, primer design is efficient and cost-effective, as high-level multiplexing is possible using the same primers for PCR and SBE (a clever strategy to simplify assay design).

SNP arrays were developed for the express purpose of genotyping individuals. However, advances in data processing and analysis have now allowed the same arrays to be used to detect both gene and chromosome (aneuploidy) copy number variation (Mao et al., 2007).

A new class of signatures that has shown clinical relevance is epigenetic modification such as DNA methylation, covalent modification of histones, and microRNAs or miRNAs that serve as transcription regulators (Esteller, 2008; Portela and Esteller, 2010). In particular, miRNAs have become important as biomarkers for tumor development, progression, and chemosensitivity (Volinia et al., 2006; Wang et al., 2011). Therefore many microarray manufacturing companies have started developing chips for miRNA detection (Wang and Xi, 2013). A challenge with the detection of miRNAs is the relative low abundance in samples compared to genomic DNA or mRNA targets. While amplification of the target is a standard way to increase abundance, the size of miRNAs (20−22 base pairs) makes PCR more challenging. Zhou et al. (2011) used silica nanoparticles (SiNPs) to concentrate the miRNAs and detection by surface plasmon resonance. miRNAs from a test sample are first captured in suspension on SiNPs, allowing for better washing prior to hybridization onto microarrays. The miRNA-capturing SiNPs are then captured onto DNA microarrays through DNA−DNA hybridization. Detection of miRNAs occurs using nanoparticle-enhanced surface plasmon resonance imaging. DNA methylation patterns have also been shown to be prognostic for various cancers, and while there may not be a specific oligonucleotide probe for methylated DNA, as there is no change in sequence, microarrays can be used to capture methylated DNA following enrichment procedures like immunoprecipitation and quantify relative abundance when compared to reference materials using the standard CGH approach (Weng et al., 2009).

Data processing and analysis for microarrays is fairly complex, and variations in the production of labeled targets, hybridization protocols (e.g., wash steps hybridization times), capture probe uniformity, probe length and composition, and detection formats (e.g., Streptavidin-Phycoerythrin detection versus Cy3 detection) all play into the resulting signal measured on the microarray. In addition, diverse methods of data analysis exist, and while many companies may provide software packages for commercially available microarrays, scientific design of controls into each test are key to accurate data interpretation. Dyrskjøt et al. (2010) provide a good description of the various methods of data processing and data analysis and therefore this will not be repeated in this chapter. Making scientific data open source and accessible to both researchers and developers alike has encouraged the development of open access databases for microarrays such as the Gene Expression Omnibus from the National Center for Biotechnology Information and Array Express from the European Bioinformatics Institute. These databases follow standards developed by the Functional Genomics Data Society (FGED) through the Minimum Information About a Microarray Experiment (MIAME; Brazma et al., 2001).

The six most critical elements contributing toward MIAME are:

the raw data for each hybridization (e.g., Affymetrix raw data file format, Genpix result files);

the final processed (normalized) data for the set of hybridizations in the experiment (study) (e.g., the gene expression data matrix used to draw the conclusions from the study);

the essential sample annotation, including experimental factors and their values (e.g., compound and dose in a dose-response experiment);

the experimental design, including sample data relationships (e.g., which raw data file relates to which sample, which hybridizations are technical, which are biological replicates);

sufficient annotation of the array (e.g., gene identifiers, genomic coordinates, probe oligonucleotide sequences, or reference commercial array catalog number); and

the essential laboratory and data processing protocols (e.g., what normalization method has been used to obtain the final processed data).

The International Standards for Cytogenomic Arrays (ICSA) is a consortium of clinical cytogenetics and molecular genetics laboratories that collaborate to develop protocols for the use of CMAs (as described above); make recommendations for the standards of design, resolution, and content of cytogenetic arrays; develop standards for the interpretation of CMA data; and maintain an online database for copy number variation data and disease associations to facilitate research and biomarker selection.

18.2.2 Suspension DNA Microarrays

The application of microbeads or microspheres as platforms to capture nucleic acids for detection was first introduced in the early 2000s with flow cytometry being the detection mechanism (Nolan and Sklar, 2002). Microsphere arrays are now routinely used as suspension arrays in microfluidics platforms and immobilized on solid surfaces (Rödiger et al., 2014). This section will focus primarily on suspension arrays using microspheres, as this format has different features compared to any type of solid phase array, be it covalently linked DNA probes on glass slides or covalently linked DNA probes to immobilized microspheres. Compared to standard planar DNA microarrays, suspension arrays seem to be easier to use, are often reasonably priced, and have faster hybridization kinetics (Dunbar, 2006). A significant feature is flexibility in array preparation as different types of microspheres can be used as arrays (Deshpande and White, 2012). These include isotropic microspheres that are dyed with a spectrum of fluorescent dyes, which are used in many commercially available FDA-cleared genetic tests (e.g., Cystic fibrosis testing kits drug metabolizing enzyme kits by Luminex Molecular Diagnostics, Inc.). These microspheres have a single size. The best example is the microsphere array from Luminex, which is about 5.4 μM in diameter. Arrays for analysis of up to 500 markers are available through this company. Anistropic particles can also be used to develop arrays; rather than encoding on fluorophore intensity or color, they are synthesized using polymeric, oxide, and metallic materials and have a range of sizes, shapes, and particle densities. They can be encoded not only optically but also by shape, size, and composition, thereby increasing the range of multiplex (Cederquist et al., 2010). Examples of anisotropic particle arrays include Raman dye-labeled particles, quantum dot barcodes, notched silicon particles, metal nanowires, Illumina Veracodes, oblong dot-coded particles, and magnetic barcoded hydrogel microparticles (Cederquist et al., 2010; Chen et al., 2013; He et al., 2008). Readout methods for these arrays include optical microscopy, Raman spectroscopy, and flow cytometry. Planar or solid phase microarrays need several washing steps and specialized instrumentation for signal detection, such as a laser scanner, whereas many suspension array assays can be read out with conventional laboratory equipment like a flow cytometer or a fluorescence microscope.

In terms of assay chemistry, while the capture of targets onto suspension arrays is via hybridization, various solution-based chemistries can be employed prior to the capture and readout, and therefore amplification of the targets is more feasible than what can be done with planar two-dimensional arrays. The capture of target is also performed in solution, and as the capture oligos are not spatially constrained, there is more flexibility for capturing longer PCR products. Additionally, different assay chemistries like ligation for target detection can also be performed more easily. Kipp et al. (2011) demonstrated the use of the multiplex ligation-dependent probe amplification assay for the detection of deletion markers for alpha thalassemia and evaluated its utility in clinical sample testing after implementing the assay at the Mayo clinic in Rochester, Minnesota (Kipp et al., 2011). Target detection occurred via the ligation of probes followed by the amplification of the ligated products using universal PCR primers. The amplified ligated probes were then hybridized to suspension arrays, and readout occurred on a Luminex flow cytometer. Another interesting chemistry was demonstrated by Li et al. (2014a,b) to enable the detection of unlabeled targets through the covalent linkage of suspension arrays to cleverly designed stem loop oligonucleotide probes that sterically block a biotin linked to the probes. Upon hybridization to the target (miRNA in this case) the conformational switch of the immobilized probe exposes the biotin, which can then be labeled with a Streptavidin-Phycoerythrin and detected by flow cytometry. Only specific targets would hybridize, and release

changes the conformation of the immobilized probe. Advantages of such a design are that the test sample does not need to be labeled nor is there a need for labeling with multiple dyes. This technique was demonstrated in a small multiplex assay for detecting miRNAs associated with nonsmall cell lung cancer, but the design could easily be used for detecting any nucleic acid target.

Suspension arrays can be made universal through the use of design strategies that incorporate the complement to the microsphere-arrayed probes in the probe that is performing solution-based detection of the target. Consider the example shown in Fig. 18.2.

On the left panel is the example of an array that has been designed to capture PCR product for a specific target. Multiplex PCR can be performed in solution, and microspheres are covalently coupled to oligonucleotides bearing a complementary sequence to the specific PCR product that needs to be resolved from the mix of PCR products. This design requires the development of custom suspension arrays for every multiplex PCR reaction and can drive up the costs of assays. Additionally, any substitution of probes in the existing assay would require a different microsphere–oligonucleotide combination to be included. The right panel shows the Multiplex Oligonucleotide Ligation-PCR chemistry adapted to be used with suspension arrays in the Deshpande and White lab. In this case, target detection occurs via the hybridization and ligation of specific probes that have been designed to include two modular components: a universal PCR primer complement and a complement for capture onto suspension arrays that are coupled to standard universal oligonucleotides. In this design, if the probes were to be substituted with others, the only new design would be for the probe that would be a complement for the substitute signature. The exact same suspension array can be used as long as the probe design includes the appropriate modular component for capture onto the same array. Multiplex PCR can also be designed in the same way.

Functionally, suspension arrays can be used for the quantification of gene expression, the detection of unique signatures and SNPs, and the detection of a change in copy number (gene or chromosome), much the same as solid phase arrays. However, due to the shorter lifetime of this platform, its use in clinical settings is not as routine as solid phase microarrays. The best application of suspension arrays has been for the detection of signatures amplified using multiplex PCR and captured via hybridization to the microspheres (Markou et al., 2011; Nishigaki et al., 2010). This design has been applied to not only the detection of unique DNA signatures but also the detection of the

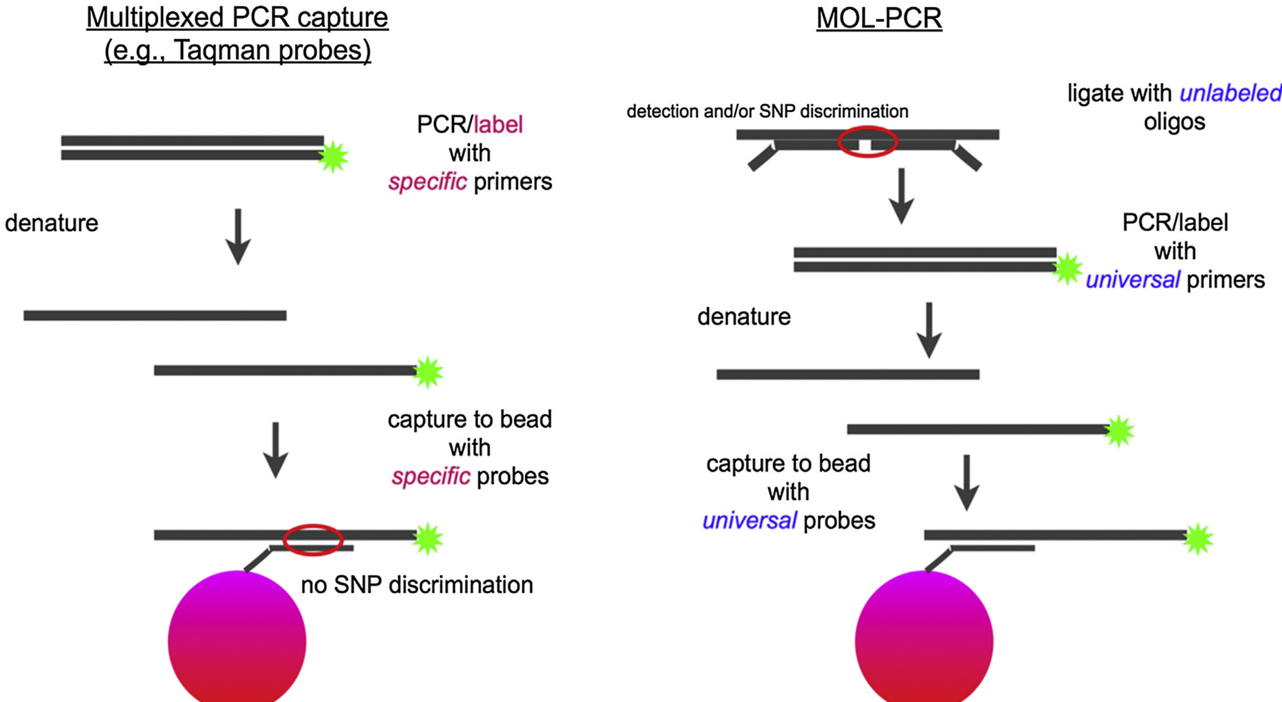


FIGURE 18.2 Use of suspension arrays to readout multiplex polymerase chain reaction (PCR)-based detections of nucleic acid signatures (left panel) and multiplex ligation PCR-based detection. In this picture, the use of TaqMan probes is not for signature detection but rather the capture of the amplified target onto the microsphere via the hybridization of the TaqMan probe to one of the amplicon strands. In essence, PCR primer pairs and TaqMan probes used in real-time PCR are used in a multiplex reaction such that the PCR primers amplify a specific target, and the TaqMan probe that would have been the detection component is used to capture the target amplicon onto the microsphere, thus effectively helping to resolve the multiplex reaction.

expression of unique prognostic biomarkers in cancer. Multiplex PCR on mRNA extracted from isolated circulating tumor cells followed by capture and readout on suspension arrays has been demonstrated as a low-cost, easy technique for studying the expression of prognostic markers (e.g., KRT19, a specific epithelial marker of prognostic significance; ERBB2, which indicates a response to therapy; TWIST1, a marker of epithelial-to-mesenchymal transition; and MAGEA3, correlation with metastasis) in small volume samples (Markou et al., 2011).

SNP detection using suspension arrays is commonly performed using the SBE technique similar to that used on solid phase microarrays. Fig. 18.3 shows the results of genotyping a sample using a simple 11-plex assay developed for the detection of TNF alpha promoter variants and applied to the study of HPV 16-associated cervical cancer susceptibility (Deshpande et al., 2005). DNA samples representing cervical cancer cases and controls from 787 Hispanic and non-Hispanic Caucasian women were screened for 11 SNPs using this assay, which included PCR of the promoter region followed by in-solution SBE in four separate reactions, one for each ddNTP. Rather than using four different suspension arrays, the same array was able to be used, because the SBE primers included separate tags that allowed for the capture of the SBE reaction and readout on the flow cytometer. Samples were also read in 96-well plate formats, allowing high-throughput processing. As can be seen from the figure, the genotype of the individual (homozygous or heterozygous) for each SNP is easily identifiable by the fluorescence intensity seen on the Y axis.

A rapid technique for the detection of aneuploidy was demonstrated by García-Herrero et al. (2014) where bacterial artificial chromosomes were covalently linked to microspheres and used to detect microdeletion syndromes normally tested by karyotyping by fluorescent in situ hybridization (FISH). There is a growing need for rapid prenatal testing, and assays using suspension array technology are being considered as a tool for a first screen followed by confirmation using karyotyping.

The level of multiplex for suspension arrays is definitely not as high as what Affymetrix or Agilent microarrays can achieve. On the other hand, due to the solution-based chemistries, the throughput of assays using suspension arrays is far higher, because all steps can be performed in 96- or 384-well plates, allowing for a larger number of samples to be tested. Thus suspension arrays may be more amenable to large-scale screening applications (population screening for a particular biomarker) than solid phase arrays. In addition, reaction kinetics are better in suspension than solid phase, as seen in a comparative analysis of cancer miRNA-detecting arrays of three types: locked nucleic acid microarrays (LNA), bead arrays, and low-density TaqMan real-time PCR arrays (Wang et al., 2011). As compared to a hybridization time of 16 h required for LNA arrays, the

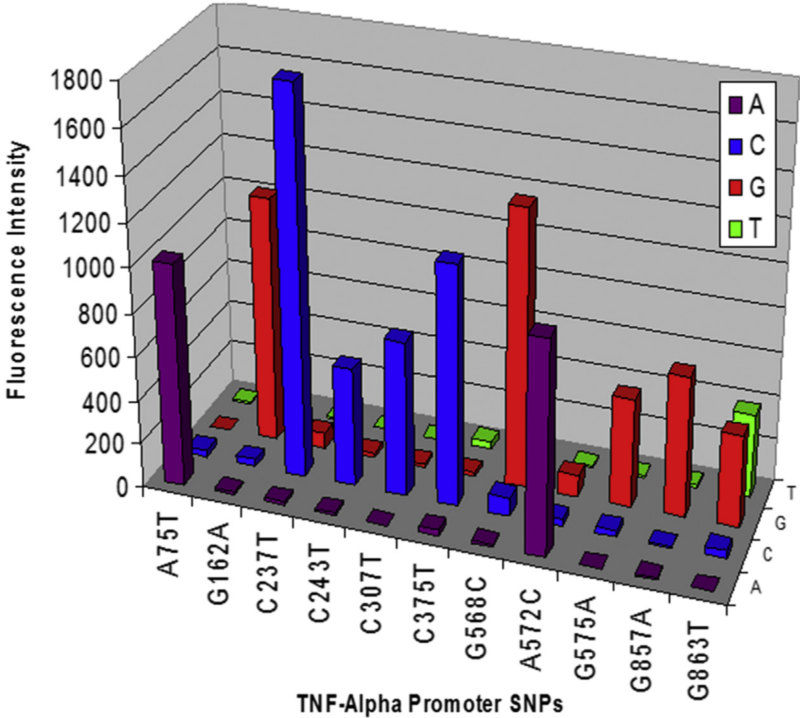


FIGURE 18.3 Demonstration of 11-plex assay to genotype a human DNA sample for TNF alpha promoter single nucleotide variants using single base extension and capture on suspension arrays.

bead arrays required just 1 h. The authors concluded that intraplatform reproducibility was acceptable for all three platforms, but interplatform reproducibility required better normalization strategies for the diverse raw data generated from the three platforms.

Table 18.1 compares and contrasts solid phase and suspension DNA microarrays. While both types of arrays have diverse applications, traditional microarrays seem more amenable to use in understanding the biology of disease and biomarker discovery, while suspension microarrays can be better applied to disease diagnosis using targeted markers, particularly in large-scale screening applications. This may be primarily due to the degree of multiplexing, which is much higher in the case of two-dimensional microarrays (but the throughput is lower). Thus two-dimensional microarrays can go deeper, and three-dimensional microarrays can be broader. There are ways to get around the multiplexing limitations through massively parallel processing, but this has to be done within practical limits.

18.3 NEW DEVELOPMENTS IN DNA MICROARRAYS AND GENETIC TESTING

Genetic testing using microarrays can be performed for disease risk estimation (Mendelian diseases like inherited cardiac disorders; Ingles et al., 2011; Hegde et al., 2008), prenatal disease detection, cancer staging, mental disorders (Rooms et al., 2011), and even pharmacogenetic testing (Yoo et al., 2009; Lagraulet, 2010; Abul-Husn et al., 2014). The field of genetic testing for personalized medicine shows a lot of promise, but it is still not universally adopted due to challenges associated with inconsistent interpretations of results, the complex interplay between genetic influence and environmental factors on drug response, and ethical considerations in regard to data sharing. "As such, the personalized medicine paradigm now includes the utilization of individual genetic data in conjunction with other clinical, family, and demographic variables to inform decisions on disease prevention, diagnosis, treatment, and prognosis" (Abul-Husn et al., 2014). Traditionally, genetic testing was controlled by the healthcare professional, and orders were sent to clinical labs on behalf of the patient. However, the concept of direct to consumer (DTC) testing that was introduced in 2007 has gained considerable momentum, with several companies offering services for genetic risk estimation for various diseases (e.g., 23andMe, deCODEme, Navigenics). It has been projected that the global DTC genetic testing market will reach greater than $230 million by 2018 (PRWeb, 2012).

Both solid phase and suspension microarrays have seen an explosion of use in genetic testing, and many companies have developed standard test kits that can be used in diagnostic labs as well as research settings. As mentioned before, microarrays have seen the most use in cancer, prenatal screening, and mental disorder diagnoses.

TABLE 18.1 Comparative Analysis of Solid and Suspension Arrays

Features	Solid Phase DNA Microarrays	Suspension DNA Microarrays
Detection chemistry	Hybridization	Hybridization
Assay chemistry	Hybridization, PCR, ligation	Hybridization, PCR, ligation
Biomarkers interrogated	Unique signatures, insertion/deletions, repeats, SNPs, copy number variations, aneuploidy	Unique signatures, insertion/deletions, repeats, SNPs, copy number variations, aneuploidy, DNA methylation
Plex level	High (thousands to millions of markers)	Moderate (hundreds)
Throughput	Low to medium (tens of samples)	High (hundreds of samples)
Stronger application	Biomarker discovery, understanding etiology of disease, quantitative	Disease diagnostics, semiquantitative
Cost for standard assay	Single kit, 5–6 samples, thousands	Single kit, up to 96 samples, thousands
Sample preparation	Standard extraction protocols for DNA and RNA	Standard extraction protocols for DNA and RNA
Assay time	Up to 48 h	Up to a single working day
Readout	Laser scanning, surface plasmon resonance	Fluorescence detection, spectroscopy, flow cytometry

PCR, polymerase chain reaction; *SNP*, single nucleotide variant.

Guidelines have been proposed for the use of microarrays in standard clinical practice for the diagnosis of chromosomal abnormalities (Manning and Hudgins, 2010), and there is a push for the use of microarrays (solid phase or suspension) as rapid screening tools for markers too small to be detected by the gold standard cytogenetic techniques such as G-banding or FISH.

While traditional descriptions of DNA microarrays only focused on solid phase detection platforms, given the expansion of assays using suspension arrays, it seems relevant to expand the definition of DNA microarrays and describe both types of platforms. The field of DNA microarrays has reached a fairly mature stage, and the use of this detection platform has become mainstream. However, there are still challenges of cost and throughput that researchers are developing ideas to address. The following paragraphs provide a glimpse into some promising platform designs that could help achieve the golden goal of high throughput and high multiplexing simultaneously.

The use of magnetic gold nanoparticles, together with readout on solid phase glass slides, is an interesting idea being worked on to improve the throughput of solid phase microarrays and reduce costs per sample. SNP arrays traditionally suffer from lower discrimination due to melting temperature differences between hybridizing probes. A solution to this is to perform SBE reactions to identify the SNP, which can also be done on a solid phase microarray. However, if performed in solution, reaction kinetics enable more efficient base extension reactions, the products of which can then be captured on suspension array magnetic gold nanoparticles (Li et al., 2012). Once captured, the suspension arrays are then immobilized for readout by spotting on low-cost glass slides. This format is conducive for applications to large-scale SNP detection on a large number of samples.

Paper-based detection platforms have been the focus of research in an effort to address the expensive instrumentation that is needed to perform readouts of DNA microarrays, whether they are solid phase or suspension. There have been quite a few applications of paper-based assays for the detection of protein biomarkers, and Song et al., (2014) have demonstrated a paper-based DNA array technology by combining cellulose probe coupling with hybridization using capillary transport. The design using paper was exactly the same as that used for suspension arrays. Standard cellulose filter paper was functionalized so that it could be covalently coupled to oligonucleotide probes via carbodiimide chemistry, akin to what is used for microspheres and oligonucleotides that capture PCR products from a multiplex PCR reaction arrayed on paper similar to that for glass slides. PCR products were labeled with paramagnetic beads as opposed to a fluorophore, and the PCR product was allowed to diffuse through the paper by capillary action. Amplified targets hybridized to the specific capture oligonucleotide on the paper, and due to the naturally brown color of the paramagnetic beads, there was a visual detection of the positive sample.

Using a clever probe-design strategy, Li et al. (2014a,b) introduced the concept of universal solid phase microarrays that can bind to DNA, miRNA, protein, and small molecules on a single slide. They developed a DNA nanostructure-based microarray platform, where the nanostructures covalently linked to glass slides were DNA tetrahedral, carrying three amino groups and one single-stranded DNA extension. The researchers used self-assembly of four oligonucleotides followed by high-performance liquid chromatography purification to develop the tetrahedral, which were then covalently coupled to glass substrates via amine–aldehyde interactions. The three amine vertices of the tetrahedral were used to immobilize the probe onto the glass surface, and the single-stranded DNA extension serves as the probe that can be differentially modified to bind nucleic acid, protein, or small molecules like cocaine. While only proof of principle has been shown here, this research demonstrates a path toward making microarray technology competitive by using low-cost substrates like aldehyde-functionalized glass and multifunctional probes. As mentioned before, universal suspension arrays are used routinely by designing modular complementary tags to the probes used in solution chemistries for detecting a target. However, the scope of such "universal" DNA arrays is restricted to coupling to different assay chemistries like PCR, hybridization, or ligation.

While next-generation sequencing (NGS) is gaining ground, and there are projections that this technology will replace all other forms of molecular analysis, there are significant challenges associated with the cost of implementation of NGS and the standardization of data processing and analysis (Pant et al., 2014). Operationally, microarray-based assays are still faster and easier to interpret, as seen in direct comparisons by researchers. Juneau et al. (2014) compared a microarray-based and an NGS-based quantification of Digital Analysis of Selected Regions (Norton et al., 2012) assay products from maternal cell-free (cf) DNA obtained from 878 maternal venous blood samples. Microarrays provided more accurate cfDNA analysis than NGS and could be performed in less time. The key underlying issue was the difference between targeted data generation versus irrelevant data that has to be sifted through just because it is generated. More is not always good.

In the realm of human genetic testing, the application of microarrays, solid phase, or suspension is limited to human genetic and epigenetic markers. However, as more research points toward the complex influences of the human microbiome on human disease (Cho and Blaser, 2012), microbiome markers may need to be included with human genetic markers in testing. Community genome arrays, functional gene arrays, and phylogenetic microarrays have already been developed and are in use to understand the

microbiome composition, function, and role in health and disease (Paliy and Agans, 2012; Tottey et al., 2013; Chen et al., 2014). It seems to be only a matter of time until microbiome biomarkers are discovered and included in the portfolio for genetic disease diagnostics.

In today's world of cost savings and process efficiency, perhaps the greatest challenge that the field of DNA microarrays faces is not what biomarkers it can be applied to but rather what would be the most informative and relevant microarrays that could be used rapidly and at a low cost. With increasing competition from other multiplexing technologies, such as real-time PCR and NGS, microarrays will need to be more focused, simpler to use, require cheaper detection instruments, and disposable. Innovations will continue to be seen in assay chemistries, data-processing algorithms, and the application of new biomarkers.

REFERENCES

Abul-Husn, N.S., Obeng, A.O., Sanderson, S.C., Gottesman, O., Scott, S.A., 2014. Implementation and utilization of genetic testing in personalized medicine. Pharmacogenomics Pers. Med. 7, 227−240.

Allison, D.B., Cui, X., Grier, P., Sabripour, M., 2006. Microarray data analysis: from disarray to consolidation and consensus. Nat. Rev. Genet. 7, 55−65. http://dx.doi.org/10.1038/nrg1749.

Bos, J.M., Towbin, J.A., Ackerman, M.J., 2009. Diagnostic, prognostic, and therapeutic implications of genetic testing for hypertrophic cardiomyopathy. J. Am. Coll. Cardiol. 54 (3), 201−211.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., et al., 2001. Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. Nat. Genet. 29 (4), 365−371.

Bumgarner, R., 2013. DNA microarrays: types, applications and their future. Curr. Protoc. Mol. Biol. 0 22 (Unit−22.1) http://dx.doi.org/10.1002/0471142727.mb2201s101.

Cederquist, K.B., Dean, S.L., Keating, C.D., 2010. Encoded anisotropic particles for multiplexed bioanalysis. Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol. 2, 578−600.

Chen, K., Leo, Y.T., Chou, F.S., Chan, W.C.W., 2013. Fabrication of metal nanoshell quantum-dot barcodes for biomolecular detection. Nano Today 8 (3), 228−234.

Chen, Y., Qin, N., Guo, J., Qian, G., Fang, D., Shi, D., Xu, M., Yang, F., He, Z., Van Nostrand, J.D., Yuan, T., Deng, Y., Zhou, J., Li, L., 2014. Functional gene arrays-based analysis of fecal microbiomes in patients with liver cirrhosis. BMC Genom. 15, 753. http://dx.doi.org/10.1186/1471-2164-15-753.

Cho, I., Blaser, M.J., 2012. The human microbiome: at the interface of health and disease. Nat. Rev. Genet. 13 (4), 260−270. http://dx.doi.org/10.1038/nrg3182.

Deshpande, A., White, P.S., 2012. Multiplexed nucleic acid-based assays for molecular diagnostics of human disease. Expert Rev. Mol. Diagn 12 (6), 645−659. http://dx.doi.org/10.1586/erm.12.60.

Deshpande, A., Nolan, J.P., White, P.S., Valdez, Y.E., Hunt, W.C., Peyton, C.L., Wheeler, C.M., 2005. *TNF-α* promoter polymorphisms and susceptibility to human papillomavirus 16-associated cervical cancer. J. Infect. Dis. 191 (6), 969−976.

Drăghici, S., 2003. Data Analysis Tools for DNA Microarrays. CRC Press.

Dunbar, S.A., 2006. Applications of Luminex® xMAP™ technology for rapid, high-throughput multiplexed nucleic acid detection. Clin. Chim. Acta 363 (1), 71−82.

Dyrskjøt, L., Dalsgaard-Sørensen, K., Stampe-Ostenfeld, M., Birkenkamp-Demtroder, K., Thorsen, K., Andersen, C.L., Kruhøffer, M., Jensen, J.L., Ørntoft, T.F., 2010. DNA microarrays and genetic testing. Mol. Diagn. 247.

Esteller, M., 2008. Epigenetics in cancer. N. Engl. J. Med. 358, 1148−1159.

FGED http://www.fged.org/.; http://fged.org/projects/miame/.

García-Herrero, S., Campos-Galindo, I., Martínez-Conejero, J.A., Serra, V., Olmo, O., Lara, C., Simón, C., Rubio, C., 2014. BACs-on-beads technology: a reliable test for rapid detection of aneuploidies and microdeletions in prenatal diagnosis. BioMed Res. Int., 590298. http://dx.doi.org/10.1155/2014/590298.

He, B., Morrow, T.J., Keating, C.D., 2008. Nanowire sensors for multiplexed detection of biomolecules. Curr. Opin. Chem. Biol. 12 (5), 522−528.

Hegde, M.R., Chin, E.L., Mulle, J.G., Okou, D.T., Warren, S.T., Zwick, M.E., 2008. Microarray-based mutation detection in the dystrophin gene. Hum. Mutat. 29, 1091−1099.

Ingles, J., Zodgekar, P.R., Yeates, L., et al., 2011. Guidelines for genetic testing of inherited cardiac disorders. Heart Lung Circ. 20, 681−687.

International Standards for Cytogenomic Arrays. http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000205.v5.p2.

Juneau, K., Bogard, P.E., Huang, S., Mohseni, M., Wang, E.T., Ryvkin, P., Kingsley, C., Struble, C.A., Oliphant, A., Zahn, J.M., 2014. Microarray-based cell-free DNA analysis improves noninvasive prenatal testing. Fetal Diagn Ther. 36, 282−286. http://dx.doi.org/10.1159/000367626.

Kennedy, G.C., Matsuzaki, H., Dong, S., Liu, W.M., Huang, J., Liu, G., Su, X., Cao, M., Chen, W., Zhang, J., Liu, W., Yang, G., Di, X., Ryder, T., He, Z., Surti, U., Phillips, M.S., Boyce-Jacino, M.T., Fodor, S.P., Jones, K.W., 2003. Large-scale genotyping of complex DNA. Nat. Biotechnol. 21, 1233−1237.

Kipp, B.R., Roellinger, S.E., Lundquist, P.A., Highsmith, W.E., Dawson, D.B., 2011. Development and clinical implementation of a combination deletion PCR and multiplex ligation-dependent probe amplification assay for detecting deletions involving the human α-globin gene cluster. J. Mol. Diagn. 13 (5), 549−557.

Krjutskov, K., Andreson, R., Magi, R., et al., 2008. Development of a single tube 640-plex genotyping method for detection of nucleic acid variations on microarrays. Nucleic Acids Res. 36, e75.

Kulasingam, V., Diamandis, E.P., 2008. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. Nat. Clin. Pract. Oncol. 5, 588−599. http://dx.doi.org/10.1038/ncponc1187.

Kurg, A., Tonisson, N., Georgiou, I., Shumaker, J., Tollett, J., Metspalu, A., 2000. Arrayed primer extension: solid-phase four-color DNA resequencing and mutation detection technology. Genet. Test. 4, 1−7.

LaFramboise, T., 2009. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. Nucleic Acids Res. gkp552.

Lagraulet, A., 2010. Current clinical and pharmaceutical applications of microarrays: from disease biomarkers discovery to automated diagnostics. J. Lab. Autom. 15 (5), 405−413. http://dx.doi.org/10.1016/j.jala.2010.06.011.

Li, S., Liu, H., Jia, Y., Deng, Y., Liming Zhang, L., Lu, Z., He, N., 2012. A novel SNPs detection method based on gold magnetic nanoparticles array and single base extension. Theranostics 2 (10), 967–975.

Li, D., Wang, Y., Lau, C., Lu, J., 2014a. xMAP Array microspheres based stem–loop structured probes as conformational switches for multiplexing detection of miRNAs. Anal. Chem. 86, 10148–10156. http://dx.doi.org/10.1021/ac501989b.

Li, Z., Zhao, B., Wang, D., Wen, Y., Liu, G., Dong, H., Song, S., Fan, C., 2014b. DNA nanostructure-based universal microarray platform for high-efficiency multiplex bioanalysis in biofluids. ACS Appl. Mater. Interfaces 6 (20), 17944–17953. http://dx.doi.org/10.1021/am5047735.

Manning, M., Hudgins, L., 2010. Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities. Genet. Med. 12 (11), 742–745.

Mao, X., Young, B.D., Lu, Y.J., 2007. The application of single nucleotide polymorphism microarrays in cancer research. Curr. Genom. 8 (4), 219.

Markou, A., Strati, A., Malamos, N., et al., 2011. Molecular characterization of circulating tumor cells in breast cancer by a liquid bead array hybridization assay. Clin. Chem. 57, 421–430.

Matsuzaki, H., Loi, H., Dong, S., Tsai, Y.Y., Fang, J., Law, J., Di, X., Liu, W.M., Yang, G., Liu, G., Huang, J., Kennedy, G.C., Ryder, T.B., Marcus, G.A., Walsh, P.S., Shriver, M.D., Puck, J.M., Jones, K.W., Mei, R., 2004. Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. Genome Res. 14, 414–425.

Milani, L., Fredriksson, M., Syvänen, A.C., 2006. Detection of alternatively spliced transcripts in leukemia cell lines by minisequencing on microarrays. Clin. Chem. 52, 2202–2211.

Miller, D.T., Adam, M.P., Aradhya, S., Biesecker, L.G., Brothman, A.R., Carter, N.P., Church, D.M., et al., 2010. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am. J. Hum. Genet. 86 (5), 749–764.

Nishigaki, Y., Ueno, H., Coku, J., Koga, Y., Fujii, T., Sahashi, K., Nakano, K., Yoneda, M., Nonaka, M., et al., 2010. Extensive screening system using suspension array technology to detect mitochondrial DNA point mutations. Mitochondrion 10, 300–308.

Nolan, J.P., Sklar, L.A., 2002. Suspension array technology: evolution of the flat-array paradigm. Trends Biotechnol. 20 (1), 9–12. http://dx.doi.org/10.1016/s0167-7799(01)01844-3.

Norton, M.E., Brar, H., Weiss, J., et al., 2012. Non-Invasive Chromosomal Evaluation (NICE) Study: results of a multicenter prospective cohort study for detection of fetal trisomy 21 and trisomy 18. Am. J. Obstet. Gynecol. 207, 137 e1–8.

Paliy, O., Agans, R., 2012. Application of phylogenetic microarrays to interrogation of human microbiota. FEMS Microbiol. Ecol. 79, 2–11.

Pant, S., Weiner, R., Marton, M.J., April 17, 2014. Navigating the rapids: the development of regulated next-generation sequencing-based clinical trial assays and companion diagnostics. Front. Oncol. http://dx.doi.org/10.3389/fonc.2014.00078.

Portela, A., Esteller, M., 2010. Epigenetic modifications and human disease. Nat. Biotechnol. 28, 1057–1068. http://dx.doi.org/10.1038/nbt.1685.

PRWEb, 2012. Future of Direct-to-Consumer (DTC) Genetic Testing Market Remains Fraught. http://www.prweb.com/releases/DTC_genetic_testing/direct_to_consumer_tests/prweb9780295.htm.

Rödiger, S., Liebsch, C., Schmidt, C., Lehmann, W., Resch-Genger, U., Schedler, U., Schierack, P., 2014. Nucleic acid detection based on the use of microbeads: a review. Microchim. Acta 181, 1151–1168. http://dx.doi.org/10.1007/s00604-014-1243-4.

Rooms, L., Vandeweyer, G., Reyniers, E., et al., 2011. Array-based MLPA to detect recurrent copy number variations in patients with idiopathic mental retardation. Am. J. Med. Genet. 155A, 343–348.

Shen, Y., Dies, K.A., Holm, I.A., Bridgemohan, C., Sobeih, M.M., Caronna, E.B., Miller, K.J., 2010. Clinical genetic testing for patients with autism spectrum disorders. Pediatrics 125 (4), e727–e735.

Song, Y., Gyarmati, P., Araujo, A.C., Lundeberg, J., Brumer, H., Ståh, P.L., 2014. Visual detection of DNA on paper chips. Anal. Chem. 86, 1575–1582. http://dx.doi.org/10.1021/ac403196b.

Stanford microarray database. http://smd.princeton.edu/resources/rescorporate.shtml.

Steemers, F.J., Chang, W., Lee, G., Barker, D.L., Shen, R., Gunderson, K.L., 2006. Whole-genome genotyping with the single-base extension assay. Nat. Methods 3, 31–33.

Tottey, W., Denonfoux, J., Jaziri, F., Parisot, N., Missaoui, M., Hill, D., et al., 2013. The human gut chip "HuGChip", an explorative phylogenetic microarray for determining gut microbiome diversity at family level. PLoS One 8 (5), e62544. http://dx.doi.org/10.1371/journal.pone.0062544.

Trevino, V., Falciani, F., Barrera-Saldaña, H.A., 2007. DNA microarrays: a powerful genomic tool for biomedical and clinical research. Mol. Med. 13 (9–10), 527–541. http://dx.doi.org/10.2119/2006-00107.

Volinia, S., Calin, G.A., Liu, C.G., Ambs, S., Cimmino, A., Petrocca, F., Visone, R., et al., 2006. A microRNA expression signature of human solid tumors defines cancer gene targets. Proc. Natl. Acad. Sci. U. S. A. 103 (7), 2257–2261.

Wang, B., Xi, Y., 2013. Challenges for microRNA microarray data analysis. Microarrays 2 (2), 34–50.

Wang, B., Howel, P., Bruheim, S., Ju, J., Owen, L.B., et al., 2011. Systematic evaluation of three microRNA profiling platforms: microarray, beads array, and quantitative real-time PCR array. PLoS One 6 (2), e17167. http://dx.doi.org/10.1371/journal.pone.0017167.

Wapner, R.J., Martin, C.L., Levy, B., Ballif, B.C., Eng, C.M., Zachary, J.M., Savage, M., et al., 2012. Chromosomal microarray versus karyotyping for prenatal diagnosis. N. Engl. J. Med. 367 (23), 2175–2184.

Weng, Y.I., Huang, T.H., Yan, P.S., 2009. Methylated DNA immunoprecipitation and microarray-based analysis: detection of DNA methylation in breast cancer cell lines. Methods Mol. Biol. 590, 165–176. http://dx.doi.org/10.1007/978-1-60327-378-7_10.

Yoo, S.M., Choi, J.H., Lee, S.Y., Yoo, N.C., 2009. Applications of DNA microarray in disease diagnostics. J. Microbiol. Biotechnol. 19 (7), 635–646.

Zhou, W.J., Chen, Y., Corn, R.M., 2011. Ultrasensitive microarray detection of short RNA sequences with enzymatically modified nanoparticles and surface plasmon resonance imaging measurements. Anal. Chem. 83, 3897–3902.

Chapter 19

Bioinformatics Tools for Data Analysis

L. Koumakis[1], C. Mizzi[2] and G. Potamias[1]
[1]*Foundation for Research and Technology – Hellas (FORTH), Heraklion, Greece;* [2]*University of Malta, Msida, Malta*

19.1 INTRODUCTION

High-throughput technologies for gene-expression and sequencing assessment [microarray or next-generation sequencing (NGS)] have vastly improved the ability of life scientists to detect and quantify sequences as well as the expression of genes, proteins, and metabolites. A vast number of bioinformatics tools have been developed to analyze and extract information from the experimental data. At the same time, systems biology studies the behavior of biological components such as molecules, cells, organisms, or species "as a whole." The primary aim of systems biology is to use and discover computational models of interacting genes, proteins, and cells aiming to "discover" and "reproduce" the underlying molecular machinery. Gene regulatory networks (GRNs) are the part of systems biology that deals with the modeling of gene interactions in a cell. These models have been developed to mathematically model GRNs. Most GRNs are based on laboratory experimental observations, which make their generation, validation, and curation very difficult and time-consuming tasks.

An important requirement for biologists is the need to associate gene-expression measurements with GRN regulatory machinery in order to get biologically relevant insights. Using information from GRNs in gene-expression data analyses, biomedical researchers aim to extract more accurate and meaningful results. In a general setting, given a certain GRN or part of it (a subnetwork), a particular gene selection process could focus (or "prioritize") just on the genes participating in the network. On the other side, the systems biology community takes advantage of the decoding and functional annotation of the human genome and high-throughput technology to automatically reconstruct and validate GRNs. The fundamental underlying assumption of the approach is that strong associations of genes in gene-expression profiles are indications for respective gene-to-gene (even distant) interactions in a GRN.

19.2 NEXT-GENERATION SEQUENCING PIPELINES

The process required for extracting information from an NGS experiment can be divided into three processes: primary analysis, whereby the raw sequence reads are obtained from the sequencer; secondary analysis, whereby the ported NGS reads are assembled with the human reference genome; and tertiary analysis, which produces an interpretation of the results of the alignment.

The output from current NGS instruments consists of sequenced reads containing nucleotides independent of the experiment type or the machine used. A number of bioinformatics tools exist for the remaining analyses, which focus on preprocessing, assembling, annotation, and data interpretation bioinformatics tools used to analyze NGS data depending on the experiment [RNA-Sequencing (RNA-Seq), targeted or genome sequencing, and de novo sequencing]. A variety of software applications have been developed to tackle a number of NGS shortcomings Before sequence reads from NGS platforms can be assembled to a target reference sequence, the data must be "preprocessed" to "clean" the data for subsequent analyses. Preprocessing may involve filtering out low-quality reads, changing the file formats, generating statistics on produced reads or nucleotide changes, or trimming the ends of reads. Various software tools and packages can be used for data preprocessing, including the Genome Analysis Toolkit (GATK) (McKenna et al., 2010) and the Fastx-Toolkit, which is implemented with other software at the Galaxy online toolkit (Goecks et al., 2010; Giardine et al., 2005; Blankenberg et al., 2001). Once the reads have been "cleaned" one needs to assemble the sequence, where reads

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00019-5
Copyright © 2017 Elsevier Ltd. All rights reserved.

are aligned to the reference genome. Using the Burrows-Wheelers Aligner (Li and Durbin, 2010) the sequenced can be aligned to the human reference genome, which is constantly updated. Once the genome has been aligned, the reads stack to what is known as coverage (the number of reads aligned to the reference sequence at a given location). Various research discusses the importance of having deeper coverage and locations where coverage is low or nonexistent due a number of factors, including sequence repeats, gaps, and read quality. Having the genome aligned, one can graphically represent the alignments and view variants (differences) from the reference genomes using visualization tools, including the Integrative Genomics Viewer (Thorvaldsdottir et al., 2013).

A variant is described as a location on the aligned sequence that has different nucleotides from the selected reference genome. NGS technologies can identify different types of variants, including insertions, deletions, single or tandem substitutions, and structural variants like copy number variants. Software used to call variants includes GATK and SAM Tools (Li et al., 2009a).

The variants called from the reads need to be annotated (i.e., described). There are various software or toolkits that are used for annotation. These include Annovar (Wang et al., 2010) through Galaxy, SnpEff (Cingolani et al., 2012). These software solutions use various algorithms and databases to give information regarding the novelty of variants using dbSNP and the 1000 Genomes Project, variants' locations, effects on the protein sequence, and other information, depending on the chosen parameters. Using an HGMD database one can also annotate variants with regard to whether they cause a disease or are associated with a disease/disorder predisposition. Different algorithms have been designed to find out if exonic variants are damaging to the protein structure or if one can predict it by Sorting Tolerant From Intolerant (Flanagan et al., 2010), Polyphen-2 (Adzhubei et al., 2010), or the UMD-Predictor (Frederic et al., 2009).

The final part in the analysis of sequenced genomes consists of finding the differences in variants between cases and controls. This analysis involves filtering techniques and statistical techniques. For example, to analyze family-based sequencing data, one should look at the open source SeqHBase toolkit. SeqHBase (He et al., 2015) allows for the detection of de novo variants found in families and highlights inherited homozygous or compound heterozygous mutations, which are found in effected samples in a given pedigree.

To extract information from RNA-Seq one can use Tuxedo protocol, which is based on the Bowtie, Tophat, and Cufflinks software. Bowtie (Langmead et al., 2009) is used as a fast short reads aligner to a reference genome. The aligner works on multiple cores and uses the Burrows-Wheeler index for low memory usage, thus it is claimed that it can "align more than 25 million reads per CPU hour with a memory footprint of approximately 1.3 gigabytes." Tophat (Trapnell et al., 2009) is used to align RNA-Seq reads with a reference genome and discovers known or novel splicing mutations. The alignment is done using Bowtie, and Tophat allows Bowtie to align reads to different locations to be able to discover novel splicing junctions. Tophat was tested on previously analyzed RNA-Seq experiments, discovering almost 20,000 unknown splice junctions. For RNA-Seq, the third analysis involves finding the differential expressed genes, which can be done using Cufflinks (Trapnell et al., 2013). Estimating the abundances of aligned reads using statistical tests and regulation in RNA-Seq finds genes that are differentially expressed between control and case data sets.

19.2.1 Next-Generation Sequencing Bioinformatics Packages

With the decrement of sequencing costs, research projects are focusing on sequencing a greater number of samples, resulting in the need for protocol standardization, high-performance computing (HPC), and storage space. A number of bioinformatics tools have been developed, including yabi (Hunter et al., 2012), which is an open source software that is adaptable to a range of both pluggable execution and data backend runnable on high-performance computers. Yabi allows the creation of analysis workflows through graphical interface or command line driven, which could be reused on different raw or processed large-scale biomedical data sets, including those generated with NGS platforms. Yabi can be administered through a web-driven console without the constant need for IT administrators.

Cloud computing is a solution for HPC and storage where data is stored and analyzed from an online server (hosted internally or on the web), commonly referred to as "in the cloud." There are a number of bioinformatics packages for NGS using cloud computing. Myrna (Langmead et al., 2010) is an open source cloud system used to analyze RNA-Seq, focusing on Bowtie and Bioconductor (an R package) for analysis. This application can be used on a local machine or on a cluster to take advantage of HPC. Galaxy is an open source web-based platform used as a bioinformatics platform for NGS analysis. It contains a set of tools for manipulation and filtering of data, mapping, RNA-Seq, and variant calling. Galaxy can be used on their freely available web cloud or set up on a local network. There are commercially available cloud-based solutions, including the Amazon Elastic Compute Cloud (http://aws.amazon.com/ec2/), which can work with the 1000 Genomes Project and others. The Beijing Genomics Institute (BGI) offers a commercial solution, BGI Online, to analyze NGS data. BGI Online promises to "provide easier and more effective solutions for a wide range of problems faced by bioinformaticians. With more robust software, a high

level of data security, and well-designed data logistics, BGI Online liberates researchers from the constraints of limited storage space, high labor costs, data delivery, and endless command lines in data analysis."

19.3 MOLECULAR PATHWAY ANALYSIS: WHY AND WHAT?

High-throughput, and in particular microarray technology, has advanced life scientists' ability not only to detect but also to quantify gene expression for targeted phenotypes. Initial expectations were that microarrays would reveal specific gene coexpression patterns (gene signatures) for various phenotypes, but the utility of gene-expression profiles seems to be bounded by a number of limitations, mainly because of the complexity and the individual variations and heterogeneities associated with the induced gene signatures (Ein-Dor et al., 2005; Iwamoto and Pusztai, 2010).

The usual computational task involving gene-expression data concerns the selection of the most discriminant genes. Due to limitations in DNA microarray technology, higher differential expressions of a gene do not necessarily reflect a greater likelihood of the gene being related to a disease, and therefore focusing only on the genes with the highest differential expressions might not be the optimal procedure. Fig. 19.1 provides an artificial, still indicative example of the limitations in analyzing solely gene-expression data. Sample cases 1, 2, and 3 are assigned to the "POS" class (phenotype in the biomedical domain) and sample cases 4 and 5 to the "NEG" class. At first sight we may observe that no sole gene or group of genes can discriminate between the two classes ("POS" and "NEG"). Initiating a decision tree inductive process could do this; all of the tree branches conclude to multiclass assignments.

Confronted with such limitations, the bioinformatics and systems biology research communities focus on more enhanced methodologies that utilize knowledge from known and established molecular pathways, especially in the form of GRNs, in an effort to combine and couple such knowledge with gene-expression data. GRNs encompass additional biological features, such as the network's topology and the underlying gene-to-gene interactions (i.e., expression activation, inhibition etc.) that may efficiently address relevant statistical barriers in gene selection (Hung et al., 2012). In particular, gene interaction knowledge solves the major problem of conflicting constrains when two significantly upregulated genes increase the enrichment of the gene set in expression data, even if the first gene inhibits the other in a GRN.

Fig. 19.2 highlights the paradigm shift from the mining of differential gene-expression patterns to the mining of GRN functional subpaths. Using the previous example, we matched the samples against known subpaths of GRNs. The first subpath, "IL-1R → TRADD," satisfies cases 1, 2, 3, and 5; the second subpath, "IL-1R → TRADD −| FLIP," satisfies cases 1, 2, and 3; the third subpath satisfies all samples; and the fourth subpath does not satisfy any case. The second subpath yields the maximum differential power, and it contains a potential function differentiation since it consists only of samples that belong to the "POS" class. In the figure, "→" represents an activation interaction, i.e., if the source gene is

genes	cases				
	POS			NEG	
	case1	case2	case3	case4	case5
IL-IR	ON	ON	ON	ON	ON
TRADO	ON	ON	ON	OFF	ON
FLIP	OFF	OFF	OFF	OFF	ON
MyD88	ON	ON	ON	ON	ON
NIK	ON	OFF	OFF	ON	OFF

FIGURE 19.1 Gene-expression data example; rows represent the genes and columns represent the cases, preassigned in two categories: POS and NEG. *ON*, upregulated gene; *OFF*, downregulated gene.

	cases				
	POS			NEG	
sub-paths	case1	case2	case3	case4	case5
IL-IR --> TRADO	ON	ON	ON	OFF	ON
IL-IR --> TRADO --| FLIP	ON	ON	ON	OFF	OFF
IL-IR --> MyD88	ON	ON	ON	ON	ON
IL-IR --> MyD88 --> NIK	ON	OFF	OFF	ON	OFF

FIGURE 19.2 Matching functional subpaths and gene-expression profiles.

"ON" (upregulated) then the target gene is also "ON," and "−|" represents an inhibition, i.e., if the source gene is "ON" (upregulated) then the target gene should be "OFF" (downregulated). Note that the inhibition relation is considered to hold for the following regulation cases: "ON" −| "OFF" or "OFF" −| "ON." The regulatory fingerprint reflected by this subpath could be considered to cause and in a way to "govern" the specific expression status of the genes.

Barabási et al. (2011) stated: "*Given the functional interdependencies between the molecular components in a human cell, a disease is rarely a consequence of an abnormality in a single gene, but reflects the perturbations of the complex intracellular and intercellular network that links tissue and organ systems.*" The authors concluded that there is progress toward a reliable network-based approach to disease modeling, but this progress is currently limited by the incompleteness of the available interactome map (i.e., the whole set of molecular interactions in a particular cell). Knowledge of GRNs, as it relates to specific phenotypes, necessarily implies that a key molecular target should be considered within the framework of its network. A network focus enables us to more effectively infer key transcriptional changes related to the specific phenotype by examining multiple downstream (or cross-talk) effectors of the target (Nam et al., 2014).

Probably the most important limitation is that the knowledge embedded in GRNs with regard to gene interactions is largely unexploited. The very purpose of the pathway diagrams is to capture our current knowledge of how genes interact and regulate each other. However, most of the existing molecular pathway analysis approaches consider only the sets of genes involved in these pathways, without taking into consideration their topology (or interactions). Moreover, some genes may exhibit multiple functions and may be involved in several pathways but with different roles. The state-of-the-art gene set analysis (GSA) pathway and its respective tools mainly utilize overrepresentation analysis (ORA) approaches, which report the enrichment of functional groups (for example, gene sets) for the genes of interest. GSA methodologies compromise connectivity in favor of computational simplicity and base their analysis on cellular components (genes, proteins, etc.), not on their connectivity (topology and interaction types) (Huang et al., 2008). Most GSA tools use the expression changes measured in high-throughput experiments only to identify pathways with unexpectedly high numbers of differentially expressed genes using ORA approaches or pathways whose genes are clustered in the ranked list of differentially expressed genes, not to directly estimate the impact of such changes on specific pathways (Judeh et al., 2013). So, GSA/ORA techniques cannot distinguish subsets of genes that are differentially expressed just above the detection threshold from genes that change by many orders of magnitude.

In an effort to accumulate the most relevant methodologies and systems in the field of molecular pathway analysis, relevant publications were screened with a focus on the annual Web Server issue of *Nucleic Acids Research* (http://nar.oxfordjournals.org/content/41/W1.toc#WebServices) and the *BMC Systems Biology* (Software articles) (http://www.biomedcentral.com/bmcsystbiol). As various systems take advantage of the Cytoscape (http://www.cytoscape.org) platform to visualize and analyze GRNs,

we also searched all of the Cytoscape plugins in order to identify more tools/applications related to the identification and assessment of discriminant pathways. We came up with more than 100 unique citations after removing duplicates from the combined searches (the screening of the two journals and the screening of the related Cytoscape plugins).

Most of the citations fall into the GSA and the GRN reconstruction approaches. Out of these citations, 48 are related to gene set enrichment analysis (GSEA), 54 are related to GRN reconstruction, and 25 are related to the identification of pathways (or subpaths) that differentiate between different phenotypes. The focus was on methods that aim toward the identification of phenotype-differentiating GRNs. So, the citations from the first two categories were rejected with the final pool of methodologies to cover 25 citations. Such a distribution is expected, since GSA and GRN reconstruction methodologies represent the earliest line of research aimed at in coupling gene-expression data and GRNs. As GSA and GRN reconstruction methods are out of the scope of this chapter, we refer the interested reader to related literature reviews (Heckera et al., 2009; Hung et al., 2012). Fig. 19.3 illustrates the main areas that combine gene-expression data and GRN knowledge, including the underlying network topology and gene−gene interactions.

19.3.1 Gene Set Analysis

GSA, also called pathway inference, is a widely followed strategy for gene-expression data analysis based on pathway knowledge. GSA focuses on sets of related genes and has established major advantages over individual gene analysis, including greater robustness, sensitivity, and biological relevance. GSA methods are able to detect biologically relevant signals and give more coherent results across different studies. GSA incorporates prior knowledge of biological pathways and other experimental results in the form of gene sets.

Most GSA methodologies treat GRNs as groups (plain list) of associated genes in order to identify the most discriminant ones within gene-expression data (Fig. 19.4, left part). Although pathways carry important information about the respective networks' structures and the correlation between genes, available GSA methods do not fully exploit such information. More and more methods are taking advantage of the topology and the underlying regulatory machinery recorded in GRNs, utilizing a variety of approaches (e.g., graph theory and network visualization). Most of the visualization toolkits display the differentiating genes on targeted GRNs (Fig. 19.4, middle). To date, none of the published GSA methodologies take advantage of nor visualize the signaling information within the networks, such as the topology and the type of associations (activation, expression, inhibition, etc.), which may exist between genes and which could prove to be essential for the reliability and interpretation of the analysis results (Fig. 19.4, right part).

Siu et al. (2009) demonstrated that correlations among pathway genes are valuable and cannot be ignored in gene-expression analysis. They introduced a methodology based on three statistical algorithms that are able to combine dependent p-values of genes within a pathway. Wang and

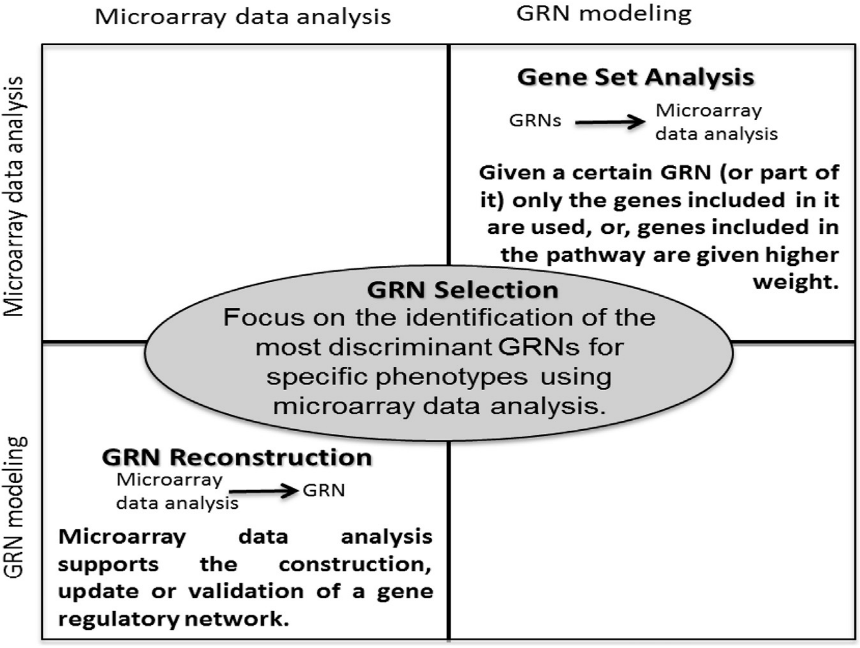


FIGURE 19.3 Scientific areas that combine gene regulatory networks and gene-expression data.

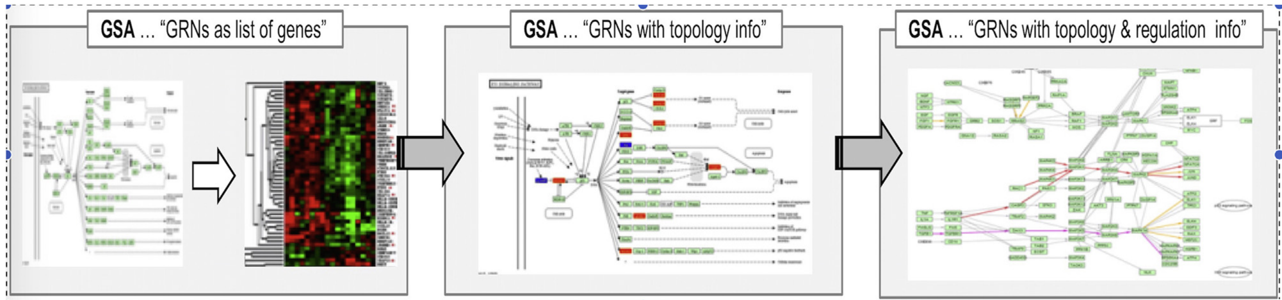


FIGURE 19.4 Evolution of gene set analysis methodologies.

coworkers (Wang et al., 2008) used a linear mixed models approach and showed that differential expression between two groups of samples is significantly different for genes in the pathway compared with the rest of the genes. They utilized pathways as a list of genes, without taking into account the underlying reactions with the pathway, under the following assumption: a pair of genes that belongs to the same pathway should be correlated.

One common approach to combine gene-expression pathway data is to incorporate known pathway information to reduce the dimensionality of gene interactions. Braun et al. (2008) proposed a method that identifies gene−pathway pairs that are highly discriminant in gene-expression data. They define the expression of a known pathway [from the Kyoto Encyclopedia of Genes and Genomes (KEGG)-pathway database] via a summary value computed following a principal component analysis approach. Tai and Pan (2007) proposed several versions of gene-expression linear discriminant analysis, taking advantage of the functional groups (i.e., pathways) to which a set of genes of interest belong. The basic underlying assumption of the proposed methods is that the genes within the same pathway will be correlated to each other. Comparative results (on simulated and real data sets) showed that these methods performed better compared to other known linear discriminant gene-expression analysis approaches.

Sfakianakis et al. (2010) proposed a method that integrates gene annotations and pathways in order to guide the cluster analysis of gene-expression profiles. The method gets information from Gene Ontology/GO (geneontology.org) to create a covariance matrix that reflects the involvement of genes in the target pathways. Utilizing an Expectation-Maximization algorithm, the maximum likelihood hidden gene variables are identified.

Eu.Gene (Cavalieri et al., 2007) is an application that tries to identify biological pathways that are transcriptionally affected under different experimental conditions. The application uses multiple pathway databases and converts them to a common format, Ensembl Gene and Transcript IDs (www.ensembl.org). The Eu.Gene Analyzer implements two different statistical methods to determine which pathways are most affected by observed gene-expression changes: the one-tailed Fisher Exact Test and GSEA.

19.3.2 Gene Set Analysis With Gene Regulatory Networks Topology Information

The web-based KEGG Mapper tool "Search&Color" (http://www.genome.jp/kegg/tool/map_pathway2.html) presents a simple tool to visualize the topology information of GRNs. The user can set a color to any gene within a molecular pathway (Fig. 19.5). A deviation from the normal regulatory network topology may reveal the mechanism of pathogenesis (Hood et al., 2004), and the genes that undergo the most changes in the network topology (as may be identified by the different colors) may serve as biomarkers for targeted phenotypes. Conversely, if several genes are involved in a pathway, but they only appear somewhere downstream, changes in their expression levels may not affect the given pathway as much.

A wealth of web-based or stand-alone toolkits take advantage of software platforms that offer visualization of complex networks. Most of the approaches rely on the Cytoscape (http://www.cytoscape.org/) toolbox.

Genoscape (Clément-Ziza et al., 2009) is an open source Cytoscape plugin that visually integrates gene-expression data from GenoScript (http://genoscript.pasteur.fr/cgi-bin/WebObjects/GenoScript) and KEGG pathways into Cytoscape networks. Genoscape automatically maps most gene or gene product identifiers to KEGG identifiers (identifiers being unique keys), enabling the import of expression data from various sources. When importing KEGG pathways, elements are filtered in order to keep only those nodes corresponding to genes or enzymes. Genoscape generates a visualization style that highlights gene-expression changes and their statistical significance. The nodes represent genes and are colored with a classical red/green gradient, according to the expression ratio level. The size of the nodes is enlarged if the corresponding expression ratio is labeled as statistically significant.

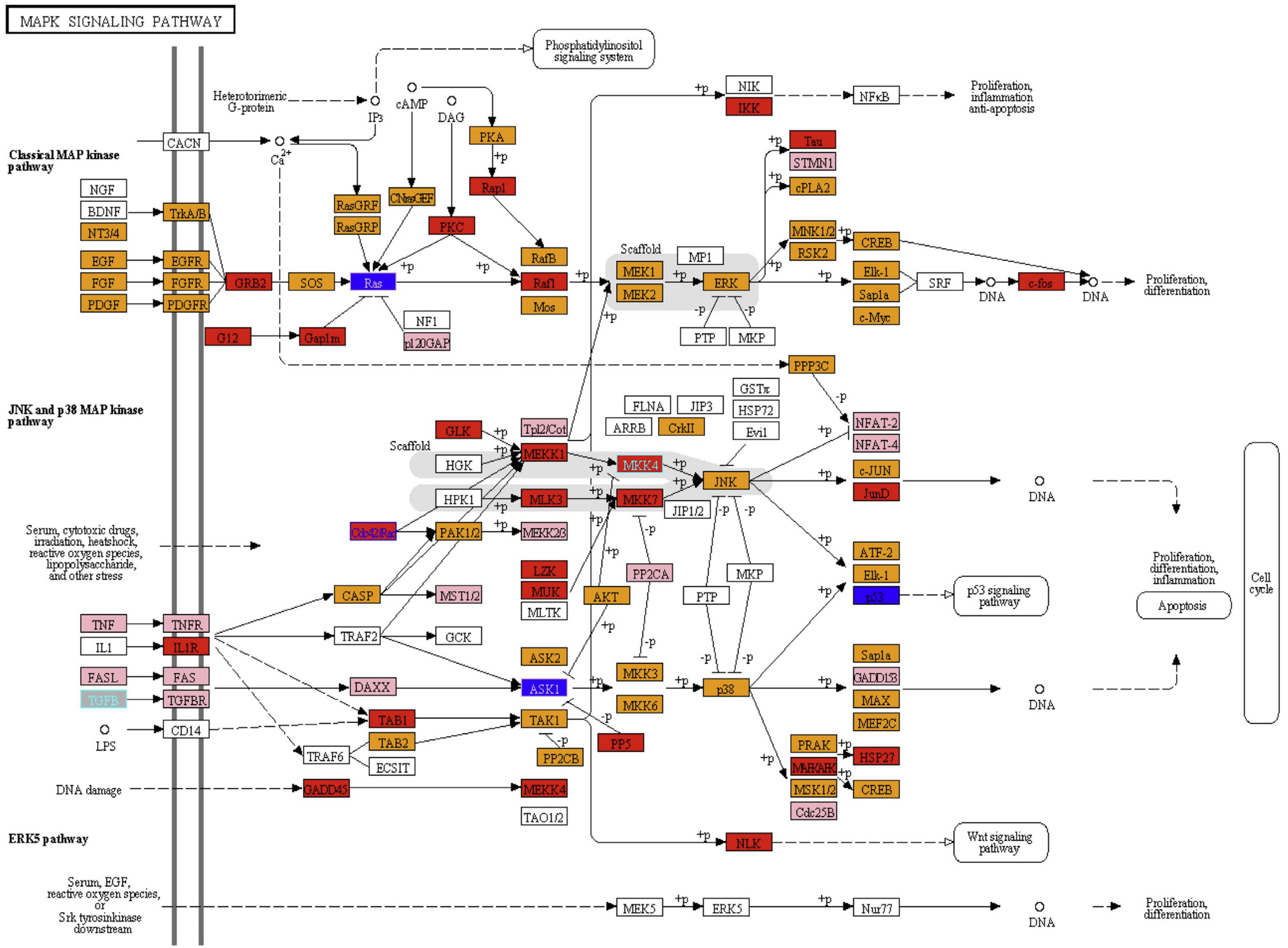


FIGURE 19.5 Example of the Kyoto Encyclopedia of Genes and Genomes Mapper/ “Search&Color” web application.

Cline et al. (2007) proposed a protocol that explains how to utilize Cytoscape to analyze the results of mRNA expression profiling and other functional genomics and proteomics experiments in the context of an interaction network. Five major steps are described: (1) obtaining a gene or protein network, (2) displaying the network using layout algorithms, (3) integrating with gene-expression and other functional attributes, (4) identifying putative complexes and functional modules, and (5) identifying enriched Gene Ontology annotations in the network.

The cancer Biomedical Informatics Grid (https://biospecimens.cancer.gov/relatedinitiatives/overview/caBig.asp) project introduced the Differential Dependency Network (DDN) (www.cbil.ece.vt.edu/software.htm) approach (Bai Zhang et al., 2011). DDN is an analytical tool for detecting and visualizing statistically significant topological changes in transcriptional networks representing two biological conditions. DDN enables differential network analysis and provides an alternative way for defining network-based predictive biomarkers. DDN has been implemented as a standalone Java application, but a Cytoscape plugin (CytoDDN) is also offered.

Ibrahim et al. (2011) described a gene-selection method that identifies groups of strongly correlated genes, which can be used to differentiate between different disease traits. In addition to utilizing static predefined pathways knowledge, the method is adaptive in the sense that it involves a pathway ranking process to identify the most relevant pathways perturbed in a given pathological state and pathway topology.

A different topological approach that utilizes various graph theory metrics is implemented within the Topology-Based Gene Set Analysis (http://www.topogsa.net/) web application (Glaab et al., 2010). The network topological properties taken into consideration are as follows: (1) the degree of a node (gene or protein), i.e., the average number of edges (interactions) incident to this node; (2) the local clustering coefficient that quantifies the probability that the neighbors of a node are connected; (3) the shortest path length for two nodes; (4) the "betweenness" of a node, calculated from the number of shortest paths; and (5) the centrality scores, computed by the entries of the dominant eigenvector (a vector that does not change when applying linear transformation) of the network adjacency matrix.

19.3.3 Identification of Phenotype-Differentiating Pathways and Subpathways

One line of research in the area of molecular pathway analysis targets the identification of those GRNs (or parts of them) that differentiate between specific phenotypes.

Draghici et al. (2003) developed Onto-Express, a tool that automatically translates lists of differentially regulated genes into functional profiles. Onto-Express introduces a methodology that identifies those pathways that contain the most discriminant genes.

Oncomine (Rhodes et al., 2007) is an advanced bioinformatics application for the identification and assessment of cancer signatures. Version 3 of Oncomine includes pathway information (from KEGG and Biocarta, http://cgap.nci.nih.gov/Pathways/BioCarta_Pathways) and respective gene-enrichment analysis functionality. The extracted signature from multiple cancer gene-expression profiles reveals pathways that are coordinately overexpressed in the respective cancer types.

Adewale et al. (2008) proposed a statistical method to identify pathways associated with targeted phenotypes (e.g., time to death for breast cancer). This is accomplished by combining the regression-based test statistic for each individual gene in a pathway of interest into a pathway-level test statistic, making the respective pathways candidates for phenotype differentiation. Then the candidate pathways are tested to determine the statistical strength of any putative association with the targeted phenotypes, with the subsequent selection of the most significant ones.

Ma and Kosorok (2010) proposed a methodology for the identification of those pathway genes that exhibit high predictive power with regard to the prognosis of breast cancer. The method is based on two quality controls: the computation of the predictive power of each gene within each pathway and the computation of the predictive power of each pathway in multiple data sets.

The Path Basic Local Alignment Search Tool (BLAST) (Kelley et al., 2004) is a tool that identifies and visually promotes pathway alignments of different networks. With PathBLAST, the user specifies a short protein interaction path as a query against a target protein–protein interaction network (selected from a network database). PathBLAST returns a ranked list of matching paths from the targeted networks along with a graphical view of these paths and their overlaps. PathBLAST performs alignment of protein networks just as BLAST is used to perform rapid alignments of proteins.

A tool for biological network integration is the GeneMANIA (Warde-Farley et al., 2010). GeneMANIA constructs and visualizes an interactive functional association network, constructed from a user-defined list of genes. Data sources used for gene similarity searches include coexpression data from Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo); physical and genetic interaction data from the Biological General Repository for Interaction Datasets (thebiogrid.org); predicted protein interaction data-based orthology information from the Interologus Interaction Database (ophid.utoronto.ca/ophidv2.204/ppi.jsp); and pathway or molecular interaction data from Pathway Commons (www.pathwaycommons.org). GeneMANIA offers also a Cytoscape plugin (Montojo et al., 2010).

An approach for the identification of differentially expressed pathways has been proposed by Nacu et al. (2007). The proposed methodology computes a score that measures to what extent a group of genes is differentially expressed. Two scoring approaches have been developed

and evaluated: target just a limited list of predefined groups and select the ones with high scores, or search for high-scoring gene sets subject to some structural constraints.

All of the methods mentioned above handle GRNs only as a group/list of genes. Information about the network topology and the underlying regulations between genes is ignored.

19.3.4 Identification of Phenotype-Differentiating Pathways and Subpathways With Gene Regulatory Network Topology Information

Chen et al. (2011) proposed a subpathway enrichment approach for the identification of a drug response principal network that takes into consideration the intrinsic pathway quantitative structures. The method is based on a fundamental assumption, namely, a subpathway may respond more effectively or with greater sensitively than the whole pathway. The methodology is based on the generation of a large number of relative subpathways (from KEGG), mapping of gene-expression data onto these pathways accompanied by their statistical scoring, and the identification of the principal components of the most perturbed subpathways. The principal components of subpathways are combined into a larger drug response network on which both topological and biological analyses are performed. The method utilizes the NetworkAnalyzer (Assenov et al., 2008) (med.bioinf.mpi-inf.mpg.de/netanalyzer/index.php) for the analysis of the topological properties of the subpathways. The NetworkAnalyzer computes and visualizes a comprehensive set of topological parameters, such as the network diameter (greatest distance of all shortest paths); the average clustering coefficients (ratio of connected edges); and shortest path lengths but ignores the underlying network regulatory mechanisms of the pathways.

The De Novo Discovery of Dysregulated Pathways in Human Diseases (DEGAS) methodology (Ulitsky et al., 2010) identifies connected gene subnetworks that are significantly enriched for those genes that are dysregulated in targeted disease specimens. DEGAS is available as part of the MATISSE (acgt.cs.tau.ac.il/matisse) software package.

Another similar approach, which uses the same algorithm for the identification of dysregulated genes/cases, is the KeyPathwayMiner (keypathwayminer.mpi-inf.mpg.de) (Alcaraz et al., 2011). Given a biological network and a set of case–control studies, KeyPathwayMiner efficiently extracts all maximal connected subnetworks. These subnetworks contain the genes that are mainly dysregulated and are differentially expressed. The exact quantities for “mainly” and “most” are modeled with two easy-to-interpret parameters that allow the user to control the number of outliers (not dysregulated genes/cases) in the solutions. KeyPathwayMiner utilizes the Cytoscape visualization library to map the dysregulated subnetworks. Version 2.0 of KeyPathwayMiner (Alcaraz et al., 2012) provides two more algorithms to solve the involved graph problems and offers an improved user interface.

Ideker et al. (2002) used subgraph extraction as a technique to predict pathways from biological networks and a set of genes. The method is more commonly used for the extraction of more complex nonlinear subnetworks in protein–protein and protein–DNA networks given yeast gene-expression data. One work by the same team applied a protein network-based approach in order to identify markers not as individual genes but as subnetworks extracted from protein interaction databases (Chuang et al., 2007). The resulting subnetworks provide models of the molecular mechanisms underlying metastasis. It is shown that the identified subnetworks are significantly more reproducible between different breast cancer cohorts than individual marker genes selected without network information. To integrate the gene-expression and network data sets, the method overlays the expression value of each gene on its corresponding protein in the network and searches for subnetworks whose activities across the patient cases are highly discriminant for metastasis. Subnetworks do not take into account relations between genes; instead, the genes are considered to be active whenever they involve highly expressed genes. These subnetworks are identified by a process of sampling the space of possible subnetworks and following a simulated annealing approach.

Wu and Stein (2012) present a semisupervised algorithm that first discovers modules of interacting genes (subpathways) involved in the disease process, independently of clinical status, and then identifies clinically significant modules using supervised principal component analysis. The implementation is based on top of a human protein functional interaction network constructed by combining curated (cleaned and normalized) and noncurated data sources. This functional interaction network covers roughly half of annotated human proteins and is highly reliable based on a variety of metrics, including confirmation of its predictions by domain experts. The network as a whole is unweighted (not scored) without regulation mechanisms.

The CLIPPER system (Martini et al., 2013) (comes as an R/Bioconductor package, www.bioconductor.org/packages/release/bioc/html/clipper.html) implements a two-step empirical approach. In the first step CLIPPER selects significant pathways according to statistical tests on the means and the concentration matrices of the graphs derived from pathway topologies. Then it “clips” the whole pathway, identifying the signal paths having the greatest association with a specific phenotype. For example, a proportional increase of the expression of the genes A and B in one of two conditions will result in a significantly different mean expression between the two conditions. The

correlation strength between A and B, however, does not change. In this case, we would have pathways with significantly altered mean expression levels but unaltered biological interactions. CLIPPER searches for pathways strongly involved in a biological process with the restriction that the mean or the variance of the expression levels are significantly altered between two phenotype conditions. CLIPPER empirically identifies the portions of the network most associated with the phenotype using the structure of the junction tree as a backbone. Even though CLIPPER uses parts of the pathway (subpathway) as a junction tree, the subpathway selection methodology ignores the relations/regulations between genes, which can be found in the signaling pathways.

Kazmi et al. (2008) developed a metaanalysis tool for identifying functional gene regulatory paths and subpaths using information from gene-expression data. The upregulated genes that participate in pathways are highlighted on the respective GRNs. The system takes advantage of the activations between genes within the pathway and tries to identify the functional paths or propose new paths. Expression values for genes missing from the gene-expression data are also added, using a predictive algorithm.

Another relevant R-based software package is the SubpathwayMiner (Li et al., 2009b) (cran.r-project.org/src/contrib/Archive/SubpathwayMiner). It is a pathway analysis tool relative to pathway annotation and identification that exploits information about the structure of metabolic pathways. The system detects distance similarity among enzymes in each pathway and mines each subpathway in which the distance among all enzymes is no greater than a (user-defined) threshold. SubpathwayMiner converts each metabolic pathway to an undirected graph (a graph where there is no direction between the connected edges) with enzymes as nodes. Two nodes in an undirected graph are connected by an edge if there is a common compound in the enzymes' corresponding reactions. As a result, the metabolic pathway is simplified when chemical compounds are omitted from the graph. Visualization of the resulting pathways is possible by linking to the KEGG visualization tools.

The main limitation of the aforementioned molecular pathway analysis approaches is that all of the gene interactions in the respective GRNs are considered to be just graph connections, i.e., they don't take into account the exact type of the interaction (e.g., activation or inhibition).

19.3.5 Identification of Phenotype-Differentiating Pathways and Subpathways With Gene Regulatory Network Topology and Regulatory Mechanisms

Geistlinger et al. (2011) stated: "As the sign of gene expression changes and the direction of regulatory interactions are so far not taken into account, substantial features of the data are still ignored and the dynamics of the transcriptomic system are not realistically reflected. Activation and inhibition are essential regulatory mechanisms in the transcriptional machinery of the cell and are causes for up- and down-regulation of particular genes."

Geistlinger et al. introduced Gene Graph Enrichment Analysis (GGEA) (www.bio.ifi.lmu.de/en/GGEA), which takes advantage of fundamental regulation types in a novel enrichment framework for signed (positive or negative) and directed (direction of connections) GRNs, to judge whether the topology of the network is a good fit for gene-expression data. GGEA performs three essential steps. First, the gene set is mapped onto the underlying GRN, yielding an induced subnetwork. Second, each edge of the induced network is scored for consistency with the gene-expression data; the signs of the expression changes of two interaction partners are evaluated for agreement with the regulation type (activation/inhibition) of the link that connects both genes. Third, the edge consistencies are summed up over the induced network, normalized, and estimated for significance using a permutation process. The GRNs are modeled as Petri Nets (networks were edges represent transitions) with fuzzy logic (a real logical value between zero and one) features. The regulations of the GRN are required to be specified with direction and effect.

Similar to GGEA, another advanced discriminant subpathway identification R/Bioconductor package is Signaling Pathway Impact Analysis (SPIA) (Tarca et al., 2009) (www.bioconductor.org/packages/release/bioc/html/SPIA.html). SPIA combines the evidence obtained from the classical GSA approaches with a novel type of evidence that measures the actual perturbation on a given pathway under a given condition. The system introduces a global probability value, P_G (calculated for each pathway), that incorporates parameters, such as the log fold-change of the differentially expressed genes, the statistical significance of the set of pathway genes, and the topology of the pathway. P_G is a combined probability value of P_{NDE} and P_{PERT}, which can be used to rank the pathways. P_{NDE} is the probability of observing a given number of differentially expressed genes or higher just by chance, and P_{PERT} is calculated following a bootstrapping process in which both the pathway and the number of differentially expressed pathway genes are fixed.

Graphite Web (Sales et al., 2013) (graphiteweb.bio.unipd.it) is a web-based GSA tool that also exploits information about pathway topology. Graphite Web implements different molecular pathway analysis methodologies (GSA, GSEA, SPIA, and CLIPPER) on three model organisms and two pathway databases (from KEGG and Reactome) and also offers pathway visualization capabilities.

Another method that identifies intergenic relationships within enriched biologically relevant subpathways is the

Topology Enrichment Analysis Framework (TEAK) (Judeh et al., 2013) (code.google.com/p/teak). TEAK employs a novel in-house algorithm and a tailor-made Clique Percolation Method to extract linear and nonlinear KEGG subpathways using Bayesian methods. The type of regulation between the nodes (genes) is always considered to be activation (e.g., for an activation "A → B" interaction, overexpression of gene A always leads to overexpression of gene B). To rank the linear and nonlinear subpathways, TEAK first uses the Bayes Net Toolbox (code.google.com/p/bnt) to fit a context-specific Gaussian Bayesian network for each subpathway (a Gaussian Bayesian network is a Bayesian network in which all of its nodes are linear Gaussians).

19.3.6 MinePath: An Integrated Molecular Pathway Analysis System

MinePath (www.minepath.org) is a web-based system that implements a discriminant analysis methodology, which elaborates on and radically extends the identification of differentially expressed GRN submodules (Koumakis et al., 2012). Besides the network topology, the method exploits the exact gene-to-gene molecular relationships ("→" activation or expression and "—|" inhibition) and identifies differentially expressed subpathways.

The MinePath methodology unfolds as a three-phase process: (1) the target GRNs are broken down into their constituent subpaths; (2) each subpath is interpreted and transformed into its binary active state (represented in binary (0,1) format); and (3) the binary representative of each subpath is aligned and matched against the discretized binary representatives of input gene-expression samples. Finally, and based on a metric that engages the proportion of matched samples, the differential power of each subpath is computed. The highly powered subpaths are kept (based on specially computed cut-off values), with the subsequent characterization of each subpath in terms of its phenotype inclination (Koumakis et al., 2015).

The basic input (detailed information can be found on the help page of MinePath) to the system includes the following: (1) selection from stored databases or the upload of new (user-specified) gene-expression data; (2) selection of

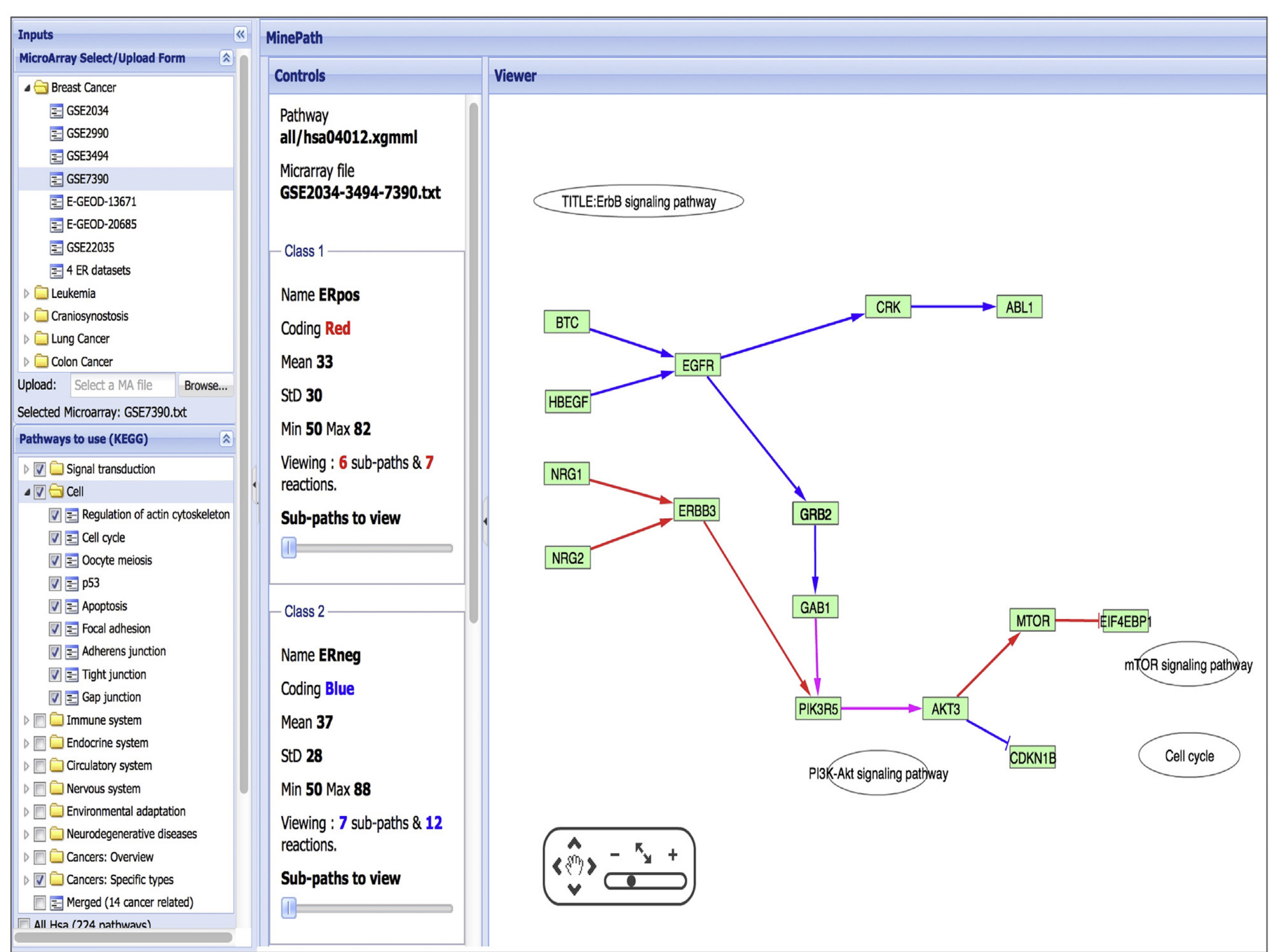


FIGURE 19.6 The MinePath web-based system (example from www.minepath.org).

target GRNs (a collection of all homo sapiens KEGG pathways is provided); and (3) optional parameterization of the subpath ranking metric along with the selection of the validation algorithm, which is seamlessly realized using the Weka API (www.weka.org) for machine learning techniques: C4.5/J48, Support Vector Machines, and NaiveBayes.

A fundamental innovation of the system concerns its advanced visualization capabilities (Fig. 19.6). All target GRNs are available for visualization (using Cytoscape Web libraries), with their KEGG topology layout being preserved. The main feature of MinePath is that it offers visualization of the active gene-to-gene regulatory mechanisms using different colors, according to their subpath phenotype inclination: red and blue for phenotype-1 and phenotype-2, respectively; magenta for active relations in both phenotypes; and orange for subpaths, which are always active. Contrary to other pathway visualization tools, MinePath calculates and enables the visualization of differentially expressed relations between genes instead of just differential genes. Furthermore, MinePath supports live interaction and immediate visualization of relations when the user sets new thresholds for the differential subpaths between the two phenotypes and the always active subpaths. The user also has the option of choosing to hide/show the overlapping relations or hide/show the associations/dissociations of the pathway. In addition, MinePath is equipped with functionalities that enable the reduction of network complexity (the deletion of genes, edge relations, and parts of the network) as well as reorientation of its topology.

19.4 CONCLUSIONS

This chapter showed how new bioinformatics pipelines were created to understand high-throughput technologies for gene-expression and sequencing assessment (microarray or NGS). With the generation of different biological data, referred to as "omics," the software and databases mentioned in this chapter made it easy to better extract knowledge out of the amount of data generated from new machines. To conclude this chapter, we firmly believe that bioinformatics is and will keep creating new software/pipelines and databases to make new discoveries from novel methodologies.

REFERENCES

Adewale, A.J., Dinu, I., Potter, J.D., Liu, Q., Yasui, Y., 2008. Pathway analysis of microarray data via regression. J. Comput. Biol. 15 (3), 269−277.

Adzhubei, I.A., Schmidt, S., Peshkin, L., et al., 2010. A method and server for predicting damaging missense mutations. Nat. Methods 7 (4), 248−249.

Alcaraz, N.M.K.H., Weile, J., Wipat, A., Baumbach, J., 2011. KeyPathwayMiner: detecting case-specific biological pathways by using expression data. Internet Math. 7 (4), 299−313.

Alcaraz, N., Friedrich, T., Kötzing, T., et al., 2012. Efficient key pathway mining: combining networks and OMICS data. Integr. Biol. 4 (7), 756−764.

Assenov, Y., Ramírez, F., Schelhorn, S.-E., Lengauer, T., Albrecht, M., 2008. Computing topological parameters of biological networks. Bioinformatics 24 (2), 282−284.

Bai Zhang, B., Tian, Y., Jin, L., et al., 2011. DDN: a caBIG analytical tool for differential network analysis. Bioinformatics 27 (7), 1036−1038.

Barabási, A.-L., Gulbahce, N., Loscalzo, J., 2011. Network medicine: a network-based approach to human disease. Nat. Rev. Genet. 12 (1), 56−68.

Blankenberg, D., Kuster, G.V., Coraor, N., et al., 2001. Galaxy: a web-based genome analysis tool for experimentalists. Curr. Protoc. Mol. Biol. John Wiley & Sons, Inc.

Braun, R., Cope, L., Parmigiani, G., 2008. Identifying differential correlation in gene/pathway combinations. BMC Bioinform. 9, 448.

Cavalieri, D., Castagnini, C., Toti, S., et al., 2007. Eu. Gene Analyzer a tool for integrating gene expression data with pathway databases. Bioinformatics 23 (19), 2631−2632.

Chen, X., X, J., Huang, B., et al., 2011. A sub-pathway-based approach for identifying drug response principal network. Bioinformatics 27 (5), 649−654.

Chuang, H.-Y., Lee, E., Liu, Y.-T., Lee, D., Ideker, T., 2007. Network-based classification of breast cancer metastasis. Mol. Syst. Biol. 3, 140.

Cingolani, P., Platts, A., Wang le, L., et al., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly 6 (2), 80−92.

Clément-Ziza, M., Malabat, C., Weber, C., et al., 2009. Genoscape: a Cytoscape plug-in to automate the retrieval and integration of gene expression data and molecular networks. Bioinformatics 25 (19), 2617−2618.

Cline, M.S., Smoot, M., Cerami, E., et al., 2007. Integration of biological networks and gene expression data using Cytoscape. Nat. Protoc. 2 (10), 2366−2382.

Draghici, S., Khatri, P., Martins, R.P., Ostermeier, G.C., Krawetz, S.A., 2003. Global functional profiling of gene expression. Genomics 81 (2), 98−104.

Ein-Dor, L., Kela, I., Getz, G., Givol, D., Domany, E., 2005. Outcome signature genes in breast cancer: is there a unique set? Bioinformatics 21 (2), 171−178.

Flanagan, S.E., Patch, A.M., Ellard, S., 2010. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. Genet. Test. Mol. Biomarkers 14 (4), 533−537.

Frederic, M.Y., Lalande, M., Boileau, C., et al., 2009. UMD-predictor, a new prediction tool for nucleotide substitution pathogenicity—application to four genes: *FBN1, FBN2, TGFBR1*, and *TGFBR2*. Hum. Mutat. 30 (6), 952−959.

Geistlinger, L., Csaba, G., Küffner, R., Mulde, N., Zimmer, R., 2011. From sets to graphs towards a realistic enrichment analysis of transcriptomic systems. Bioinformatics 27 (13), i73−i366.

Giardine, B., Riemer, C., Hardison, R.C., et al., 2005. Galaxy: a platform for interactive large-scale genome analysis. Genome Res. 15 (10), 1451−1455.

Glaab, E., Baudot, A., Krasnogor, N., Valencia, A., 2010. TopoGSA: network topological gene set analysis. Bioinformatics 26 (9), 1271−1272.

Goecks, J., Nekrutenko, A., Taylor, J., 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. 11 (8), R86.

He, M., Person, T.N., Hebbring, S.J., et al., 2015. SeqHBase: a big data toolset for family based sequencing data analysis. J. Med. Genet. 52 (4), 282–288.

Heckera, M., Lambecka, S., Toepferb, S., van Somerenc, E., Guthke, R., 2009. Gene regulatory network inference: data integration in dynamic models—a review. Biosystems 96 (1), 86–103.

Hood, L., Heath, J.R., Phelps, M.E., Lin, B., 2004. Systems biology and new technologies enable predictive and preventative medicine. Science 306 (5696), 640–643.

Huang, dW., Sherman, B.T., Lempicki, R.A., 2008. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4 (1), 44–57.

Hung, J., Yang, T., Hu, Z., Weng, Z., DeLisi, C., 2012. Gene set enrichment analysis: performance evaluation and usage guidelines. Brief. Bioinform. 13 (3), 281–291.

Hunter, A.A., Macgregor, A.B., Szabo, T.O., Wellington, C.A., Bellgard, M.I., 2012. Yabi: an online research environment for grid, high performance and cloud computing. Source Code Biol. Med. 7 (1), 1.

Ibrahim, M.-A., Jassim, S., Cawthorne, M.A., Langlands, K., 2011. A pathway-based gene selection method provides accurate disease classification. Int. J. Digit. Soc. 2 (4), 566–573.

Ideker, T., Ozier, O., Schwikowski, B., Siegel, A.F., 2002. Discovering regulatory and signalling circuits in molecular interaction networks. Bioinformatics 18 (Suppl. 1), S233–S240.

Iwamoto, T., Pusztai, L., 2010. Predicting prognosis of breast cancer with gene signatures: are we lost in a sea of data? Genome Med. 2 (11), 81.

Judeh, T., Johnson, C., Kumar, A., Zhu, D., 2013. TEAK: topology enrichment analysis framework for detecting activated biological subpathways. Nucleic Acids Res. 41 (3), 1425–1437.

Kazmi, S.A., Kim, Y.-A., Pei, B., et al., 2008. Meta analysis of microarray data using gene regulation pathways. In: 2008 IEEE International Conference on Bioinformatics and Biomedicine, pp. 37–42.

Kelley, B.P., Yuan, B., Lewitter, F., Sharan, R., Stockwell, B.R., Ideker, T., 2004. PathBLAST: a tool for alignment of protein interaction networks. Nucleic Acids Res. 32 (Web Server issue), W83–W88.

Koumakis, L., Moustakis, V., Zervakis, M., Kafetzopoulos, D., Potamias, G., 2012. Coupling regulatory networks and microarays: revealing molecular regulations of breast cancer treatment responses. In: Artificial Intelligence: Theories and Applications: 7th Hellenic Conference on AI, SETN; 2012; Lamia, Greece, pp. 239–246.

Koumakis, L., Potamias, G., Tsiknakis, M., Zervakis, M., Moustakis, V., 2015. Integrating Microarray Data and GRNs. Methods in Molecular Biology. Clifton, NJ.

Langmead, B., Trapnell, C., Pop, M., Salzberg, S., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10 (3), R25.

Langmead, B., Hansen, K., Leek, J., 2010. Cloud-scale RNA-sequencing differential expression analysis with Myrna. Genome Biol. 11 (8), R83.

Li, H., Durbin, R., 2010. Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics 26 (5), 589–595.

Li, C., Li, X., Miao, Y., et al., 2009a. SubpathwayMiner: a software package for flexible identification of pathways. Nucleic Acids Res. 37 (19), e131.

Li, H., Handsaker, B., Wysoker, A., et al., 2009b. The sequence alignment/map format and SAMtools. Bioinformatics 25 (16), 2078–2079.

Ma, S., Kosorok, M.R., 2010. Detection of gene pathways with predictive power for breast cancer prognosis. BMC Bioinform. 11, 1.

Martini, P., Sales, G., Massa, M.S., Chiogna, M., Romualdi, C., 2013. Along signal paths: an empirical gene set approach exploiting pathway topology. Nucleic Acids Res. 41 (1), e19.

McKenna, A., Hanna, M., Banks, E., et al., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1393.

Montojo, J., Zuberi, K., Rodriguez, H., et al., 2010. GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. Bioinformatics 26 (22), 2927–2928.

Nacu, S., Critchley-Thorne, R., Lee, P., Holmes, S., 2007. Gene expression network analysis and applications to immunology. Bioinformatics 23 (7), 850–858.

Nam, S., Chang, H.R., Kim, K.T., 2014. PATHOME: an algorithm for accurately detecting differentially expressed subpathways. Oncogene 33 (41), 4941–4951.

Rhodes, D., Kalyana-Sundaram, S., Mahavisno, V., Varambally, R., Yu, J., Briggs, B.B., Barrette, T.R., Anstet, M.J., Kincead-Beal, C., Kulkarni, P., Varambally, S., Ghosh, D., Chinnaiyan, A.M., 2007. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. Neoplasia 9 (2), 166–180.

Sales, G., Calura, E., Martini, P., Romualdi, C., 2013. Graphite Web: Web tool for gene set analysis exploiting pathway topology. Nucleic Acids Res. 41 (Web Server issue), W89–W97.

Sfakianakis, S., Zervakis, M., Tsiknakis, M., Kafetzopoulos, D., 2010. Integration of biological knowledge in the mixture-of-Gaussians analysis of genomic clustering. In: 10th International Conference on Information Technology and Applications in Biomedicine (ITAB), pp. 1–4.

Siu, H., Dong, H., Jin, L., Xiong, M., 2009. New statistics for testing differential expression of pathways from microarray data. Complex Sci./LNCSSITE 4, 277–285.

Tai, F., Pan, W., 2007. Incorporating prior knowledge of gene functional groups into regularized discriminant analysis of microarray data. Bioinformatics 23 (33), 3170–3177.

Tarca, A.L., Draghici, S., Khatri, P., et al., 2009. A novel signaling pathway impact analysis. Bioinformatics 25 (1), 75–82.

Thorvaldsdottir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. 14 (2), 178–192.

Trapnell, C., Pachter, L., Salzberg, S.L., 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25 (9), 1105–1111.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., Pachter, L., 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. 31 (1), 46–53.

Ulitsky, I., Krishnamurthy, A., Karp, R.M., Shamir, R., 2010. DEGAS: *de novo* Discovery of dysregulated pathways in human diseases. PLoS One 5 (10), e13367.

Wang, L., Zhang, B., Wolfinger, R.D., Chen, X., 2008. An integrated approach for the analysis of biological pathways using mixed models. PLoS Genet. 4 (7), e1000115.

Wang, K., Mingyao, L., Hakon, H., 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38 (16).

Warde-Farley, D.D.S., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C.T., Maitland, A., Mostafavi, S., Montojo, J., Shao, Q., Wright, G., Bader, G.D., Morris, Q., 2010. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 38 (Web server issue), 214–220.

Wu, G., Stein, L., 2012. A network module-based method for identifying cancer prognostic signatures. Genome Biol. 13 (12), R112.

Chapter 20

Genomic Databases: Emerging Tools for Molecular Diagnostics

G.P. Patrinos[1,2], T. Katsila[1], E. Viennas[1] and G. Tzimas[3]

[1]University of Patras School of Health Sciences, Patras, Greece; [2]Erasmus University Medical Center, Rotterdam, The Netherlands; [3]Technological Educational Institute of Western Greece, Patras, Greece

20.1 INTRODUCTION

Research into the genetic basis of human disorders has advanced in scale and sophistication, leading to very high rates of data production in many laboratories, while DNA diagnostics and electronic healthcare records are increasingly common features of modern medical practice. Therefore it should be possible to integrate all of this information in order to establish a detailed understanding of how genome sequence differences impact human health. Bioinformatics, particularly genome informatics (usually defined as informatics tools used in molecular biology), is an important scientific discipline that emerged in the postgenomic era from developments in the field of human genomics. Bioinformatics can be significantly helpful in, for example, accommodating huge data sizes (billions of base pairs); stratifying large data diversity (repeated sequences, protein domains, interspecies DNA sequence comparisons); and the integration of published references.

Major advances have been made in the characterization of genes that are involved in human diseases. Concurrently, advances in technology have led to the identification of numerous genomic variants in these genes, ranging from point mutations to large rearrangements (see previous chapters). It rapidly has become clear that the knowledge and organization of these alterations in structured repositories will be of great importance not only for diagnosis but also for clinicians and researchers. Genetic or mutation databases are referred to as online repositories of genomic variants, described for a single (locus-specific) or more (general) genes or specifically for a population or ethnic group (national/ethnic). The main applications of mutation databases are as follows: (1) to facilitate diagnosis at the DNA level and to define an optimal strategy for variant detection, (2) to provide information about variant-specific phenotypic patterns, and (3) to correlate locus-specific variant information with genome-wide features, such as repetitive elements, gene structures, interspecies conservation, mutation hotspots, recombination frequencies, and so on.

In this chapter, we will summarize the key features of the main types of genetic databases that are frequently used in molecular diagnostics, namely locus-specific and national/ethnic genetic databases. In particular, the main activities relating to these genetic database types will be highlighted, in order to describe the existing and emerging database types in this domain, emphasize their potential applications in modern medical genetics, and comment upon the key elements that are still missing and holding back the field.

20.2 HISTORICAL OVERVIEW OF GENETIC DATABASES

In 1966, with great vision, Victor McKusick made the first serious efforts toward summarizing DNA variations and their clinical consequences when he published the Mendelian Inheritance in Man (MIM), a paper compendium of information on genetic disorders and genes (McKusick, 1966). This document is now distributed electronically as the Online Mendelian Inheritance in Man (OMIM) at http://www.omim.org and is regularly updated (Amberger et al., 2015). Also, the first database collecting genomic variants from a single gene was published in 1976. It included 200 mutations from the globin gene in book format and led to the HbVar database for hemoglobin variants and thalassemia

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00020-1
Copyright © 2017 Elsevier Ltd. All rights reserved.

mutations (Hardison et al., 2002; Patrinos et al., 2004; Giardine et al., 2007b, 2014).

A decade later, David Cooper began listing mutations in genes to determine which mutational change was the most common (Cooper et al., 1998). In the mid-1990s the Human Genome Organization-Mutation Database Initiative was created by another visionary researcher, the late Richard Cotton, in order to organize “Mutation Analysis,” a new domain of genetics (Cotton et al., 1998), which evolved into the Human Genome Variation Society (HGVS; http://www.hgvs.org). Today, the stated objective of the HGVS is “...to foster discovery and characterization of genomic variations, including population distribution and phenotypic associations.”

This field is rapidly expanding, and there are several hundred databases of genomic variants available on the Internet. However, not all genetic databases fulfill quality requirements, while others have been built by researchers “on the side” for their own use. Most importantly, funding opportunities for database-related projects have always been very limited.

20.3 MODELS FOR DATABASE MANAGEMENT

In its strict definition, a database is a collection of records, each of which contains one or more fields (i.e., pieces of data) about some entity (e.g., DNA sequences, mutations) and has a regular structure that is organized in such a way that the desired information can easily be retrieved. The curator is the person (or group of persons) that is responsible for developing, updating, and ultimately maintaining a genetic database. There are mainly three types of database management models, from low to higher degrees of complexities, each with their own advantages and drawbacks.

Relational databases are the most frequently used ones, since they are very efficient for dealing with large volumes of information. A relational database is based on data organization in a series of interrelated tables. Also, information can be retrieved in an extremely flexible manner by using structured data queries. Although interest in this model was initially confined to academia, subsequently, relational databases became the dominant type for high-performance applications because of their efficiency, ease of use, and ability to perform a variety of useful tasks that had not been originally envisioned. The dominant query language for relational databases is the semistandardized structured query language (SQL), for which different SQL variants exist. The main ones are Microsoft SQL (http://www.microsoft.com/sql); MySQL (http://www.mysql.com); and postgreSQL (http://www.postgresql.org). Although critics claim that SQL is not consistent with the relational model, it works extremely well in practice, and no replacement is on the horizon. The requirement of specialized software for developing a relational database can potentially be a disadvantage, since significant computer proficiency is required.

In the very first genetic databases, information on genomic variants was provided in plain text websites, i.e., the simplest “database” format. This structure cannot be considered as a database in a strict sense, but it was often used for several database projects. The advantages for such a model were the development and maintenance simplicity, since no specific software was required. However, there were no true data querying options, apart from the standard searching tool provided by the respective Internet browser, and the database was very difficult to maintain in the case of expanded data sets. In addition, flat-file databases were the simplest database types, particularly for small-scale and simple applications. These databases had a modest querying capacity and could accommodate small to moderately big data sets. Their development required average computing skills, even though they were based on simple software. The first version of the ETHnic and National Database Operating System (ETHNOS) software (Patrinos et al., 2005a) was developed using the flat-file database model.

20.4 MUTATION DATABASE TYPES

The various depositories that fall under the banner of “mutation databases” can be categorized into three main types: General (or central) mutation databases (GMDs); locus-specific databases (LSDBs); and national/ethnic mutation databases (NEMDBs).

GMDs attempt to capture all described mutations in all genes, but each is represented in only limited detail. The included phenotype descriptions are generally quite cursory, making GMDs of little value to those wishing to understand the subtleties of phenotypic variability. GMDs tend to include only genomic variants of large effect that result in Mendelian patterns of inheritance, while sequence variations associated with no, minor, or uncertain clinical consequences are rarely cataloged. Thus GMDs provide a good overview of patterns of clinically relevant mutations and polymorphisms but almost no fine detail to aid proper understanding. The best example of a GMD is the Human Gene Mutation Database (HGMD; http://www.hgmd.org; Stenson et al., 2014), which by April 2016 contained 127,836 different records in 4860 different genes in the public release and 179,235 different records in 7189 different genes in the HGMD Professional Release (V2015.4). In this type of database there are no specific field experts maintaining them. Instead, they include published data for causative genomic variants, their distribution, and references, mostly using automated data (text) mining routines. Each causative genomic variant is entered only once in order to avoid confusion between recurrent

and identical-by-descent lesions, and the phenotypic description associated with the mutation is very limited, preventing any study on phenotypic variability. These databases are frequently referred to as a "mile wide and inch deep," as they include mutations from many genes but with a limited description. However, such databases may become increasingly useful in the postgenomic era with the advent of next-generation sequencing since they can contribute toward the understanding of novel, tentatively pathogenic variants (Karageorgos et al., 2015). A detailed survey of the available GMDs has been compiled by George et al. (2008).

In contrast to GMDs, LSDBs are concerned with just one or a few specific genes (Claustres et al., 2002; Mitropoulou et al., 2010), usually related to a single disease entity. They aim to be highly curated repositories of published and unpublished mutations within those genes, and as such, they provide a much-needed complement to the core databases. Data quality and completeness are typically high, with roughly half of the stored records pertaining to otherwise unpublished variants. The data are also very rich and informative, and the annotation of each mutant includes a full molecular and phenotypic description. Therefore these databases are referred to as an "inch wide and mile deep." For example, LSDBs will typically present each of the multiple discoveries of recurrent mutational events, allowing mutation hotspots to be identified. When these mutations occur upon different chromosomal backgrounds (linked to other mutations) such that they result in several, or different, disease features, these correlations are also recorded. A good example of an LSDB is the HbVar database (http://globin.bx.psu.edu/hbvar), initially reported by Hardison et al. (2002) and subsequently updated on a frequent basis with new features and data by Patrinos et al. (2004 and Giardine et al. (2007a, 2014). This relational database of hemoglobin variants and thalassemia mutations provides information on pathology, hematology, clinical presentation, and laboratory findings for a wealth of DNA alterations. Gene/protein variants are annotated with respect to biochemical data, analytical techniques, structure, stability, function, literature references, and qualitative and quantitative distributions in ethnic groups and geographic locations (Patrinos et al., 2004; Giardine et al., 2007a). As is common in LSDBs, entries can be accessed through summary listings or user-generated queries, which can be highly specific. A comprehensive listing of available LSDBs can be found at http://www.hgvs.org, http://www.hgmd.org, and in the literature (Cotton et al., 2007).

Finally, NEMDBs are repositories documenting the genetic composition of an ethnic group and/or population and the genetic defects (leading to various inherited disorders) and their frequencies calculated on a population-specific basis. The emergence of the NEMDB is justified from the fact that the spectrum of genomic variations observed for any gene or disease will often differ between population groups worldwide and also between distinct ethnic groups within a geographical region. Not only do NEMDBs help to elaborate the demographic history of human population groups, but they are also a prerequisite to the optimization of national DNA diagnostic services. That is, they provide essential reference information for use in the design of targeted mutation detection efforts for clinical use, and they may also serve to enhance awareness among healthcare professionals, bioscientists, patients, and the general public about the range of the most common genetic disorders (and their environmental correlates) suffered by particular population groups.

Beyond the aforementioned main database types, DNA variation is also recorded in various polymorphism databases, such as those provided at the National Center of Biotechnology Information, namely dbSNP and dbGAP (http://www.ncbi.nlm.nih.gov/projects/SNP; Agarwala et al., 2016); ClinVar (Landrum et al., 2016); and the HAPMAP Data Coordination Center (http://www.hapmap.org; The International HapMap Consortium, 2003). These resources are important in helping to complete the picture for any gene or region of interest by summarizing all of the variants that are not typically included in GMDs, LSDBs, and NEMDBs. In brief, GMDs, LSDBs, and NEMDBs share the same primary purpose of representing DNA variations that have definitive or likely phenotypic effects. They tackle this goal from very different perspectives, and there is clearly a need for these three types of resources in the various disciplines of human genetics and genomics, particularly genetic testing.

In the following paragraphs, the basic aspects of LSDBs and NEMDBs will be analyzed in detail in relation to their applications in molecular diagnostics, since these resources are more closely related to molecular genetic testing than any other genomic database type.

20.5 LOCUS-SPECIFIC DATABASES IN MOLECULAR GENETIC TESTING

In the previous chapters, it has been nicely demonstrated that the detection of DNA sequence variations can be very efficiently performed using a plethora of techniques. LSDBs can facilitate molecular diagnosis of inherited diseases in various ways. For example, LSDBs can assist in ascertaining whether a DNA variation is indeed causative, leading to a genetic disease, or benign. Similarly, some high-quality LSDBs provide detailed phenotypic information that is related to disease-causing genomic variants. Ultimately, LSDBs can assist in the selection of the optimal mutation detection strategy.

Soussi and coworkers were among the first to present LSDBs' challenges and opportunities on the basis of the p53 paradigm (Soussi et al., 2006). Laboratory data on

BRCA1 and *BRCA2* genetic variants were compared to those provided by the BIC, ClinVar, HGMD Professional, Leiden-Open (source) Variation Database (LOVD), and the Universal Mutation database (UMD; Vail et al., 2015). Findings reported a substantial disparity of variant classifications within and among the databases considered, suggesting that LSDBs' use in clinical practice is still challenging.

20.5.1 Identification of Causative Genomic Variants

If a missense variant is detected in a diagnostic laboratory, additional experiments usually need to be conducted prior to concluding that the variant in question is in fact causative in the family. This is particularly important in the case of whole-genome sequencing where many variants are reported. In the absence of a functional test, the segregation of the variant in the affected family members, the absence of this variation in a panel of at least 100 control samples, the prediction of the biochemical nature of the substitution, the region where the mutation is located, and the degree of conservation among species are some of the arguments for the causative nature of the variant. As this approach is often time-consuming, the use of an LSDB can provide researchers with valuable information to help in such a decision process.

If the variant has been reported as a causative one, its full description and the corresponding literature are provided in the LSDB. Furthermore, a comprehensive LSDB does not only include the reference sequence of the gene but also the description of structural domains and data about interspecies conservation for each protein residue.

Also, as many publications include large data sets, it is often possible to observe errors or the use of wrong variant nomenclature (see Chapter 2) due to references to an old sequence (up to 10% of errors for some publications). The use of LSDBs can be helpful to this end. Several LSDBs include an automatic nomenclature system based on a reference sequence. In other words, variant entry is done such that the variant nucleotide is automatically checked against the reference sequence at the respective variant position and named based on the official (HGVS) nomenclature (den Dunnen and Antonarakis, 2001). Interestingly, Mutalyzer, a dedicated module to automatically produce any sequence variation nomenclature (http://www.lovd.nl/mutalyzer; Wildeman et al., 2008), enables unambiguous and correct sequence variant descriptions to avoid mistakes and uncertainties that may lead to undesired errors in clinical diagnosis. Mutalyzer handles most variation types, i.e., substitutions, deletions, duplications, etc. and follows current HGVS recommendations (see Chapter 2).

Finally, a handful of LSDBs include data presentation tools to visualize their content in a graphical display. VariVis is a generic visualization toolkit that can be employed by LSDBs to generate graphical models of gene sequence with corresponding variations and their consequences (Smith and Cotton, 2008). The VariVis software package can run on any web server capable of executing Perl CGI scripts and can interface with numerous database management systems (DBMS) and even flat-file databases. The toolkit was first tested in A_1ATVar, an LSDB for *SERPINA1* gene variants, leading to α_1-antitrypsin deficiency (Zaimidou et al., 2008), and it can be integrated into generic DBMS used for LSDB development (see also paragraph 22.7).

20.5.2 Linking Genotype Information With Phenotypic Patterns

LSDBs are far more than just inert repositories, as they also include analyzing tools, which exploit computing power to answer complex queries, such as phenotypic heterogeneity and genotype/phenotype correlations. The vast majority of LSDBs, especially all LOVD-based LSDBs (see paragraph 22.7), provide phenotypic descriptions in an abstract format. In other custom-built databases, such as HbVar, phenotypic descriptions are significantly more detailed, e.g., providing information on the clinical presentation of thalassemia carriers and patients together with their hematological indices for every hemoglobin and thalassemia variant. All LSDBs attempt to enforce a controlled vocabulary to facilitate straightforward data querying. Therefore phenotype data in the vast majority of LSDBs are presented in a very basic way, such as in the form of free-text entries and/or with very little detail. There is a definite need for this situation to be improved and data content to be harmonized. There is also a general wish for the comprehensive analysis of phenotypes to occur, a goal termed "phenomics" (Gerlai, 2002; Hall, 2003; Scriver, 2004), supported by the necessary informatics solutions.

It is possible for patients with the same mutation to have a completely different phenotype. This valuable information could be useful for predictive medicine. A good example is given by the distribution of mutations of the *FBN1* gene (MIM# 134797) that are associated with Marfan syndrome (MFS) and a spectrum of conditions phenotypically related to MFS, including dominantly inherited ectopia lentis, severe neonatal MFS, and isolated typical features of MFS. MFS, the founding member of the heritable disorders of connective tissue, is a dominantly inherited condition characterized by a tall stature and skeletal deformities, dislocation of the ocular lens, and propensity to aortic dissection (Collod-Beroud and Boileau, 2002). The syndrome is characterized by a considerable variation in the clinical phenotype between families and also within the same family. Severe neonatal MFS has features of MFS and congenital contractural arachnodactyly

present at birth, along with unique features such as loose, redundant skin, cardiac malformations, and pulmonary emphysema (Collod-Beroud and Boileau, 2002). A specific pattern of mutations is observed in exons 24 to 26 in association with neonatal MFS. In fact, 73.1% of mutations are located in this region in the neonatal form of the disease, but only 4.8% of the mutations associated with a classical MFS are located in these exons (FBN1 database; http://www.umd.be; Collod et al., 1996; Collod-Beroud et al., 2003).

An interesting project attempted to connect human phenotype and clinical data in various LSDBs with data on genome sequences, evolutionary history, and function from the ENCODE project and other resources in genome browsers. Phenotypes for ENCODE (PhenCode; http://www.bx.psu.edu/phencode; Giardine et al., 2007b) is a collaborative, exploratory project to help understand the phenotypes of human mutations in the context of sequence and functional data from genome projects. The project initially focused on a few selected LSDBs, namely HbVar (*HBA2*, *HBA1*, and *HBB* genes), PAHdb (*PAH* gene; Scriver et al., 2000), etc. Interesting mutations found in a genome browser can be tracked by following links back to the LSDBs for more detailed information. Alternatively, users can start with queries on mutations or phenotypes at an LSDB and then display the results at the genome browser to view complementary information such as functional data (e.g., chromatin modifications and protein binding from the ENCODE consortium), evolutionary constraint, regulatory potential, and/or any other tracks. PhenCode provides a seamless, bidirectional connection between LSDBs and ENCODE data at genome browsers, which allows users to easily explore phenotypes associated with functional elements and look for genomic data that could explain clinical phenotypes. Therefore PhenCode is not only helpful to clinicians for diagnostics, but it also serves biomedical researchers by integrating multiple types of information and facilitating the generation of testable hypotheses to improve our understanding of both the functions of genomic DNA and the mechanisms by which it achieves those functions. These and other types of data provide new opportunities to better explain phenotypes.

20.5.3 Selection of the Optimal Mutation Detection Strategy

As LSDBs collect all published and unpublished genomic variants, they are very useful in defining an optimal genetic screening strategy, especially when targeted resequencing is needed. Therefore an overview of the distribution of variants at the exonic level can help to focus on specific exons, where most of the variants are located. The best example is given by the study of the TP53 gene involved in up to 50% of human cancers (Soussi, 2000). This gene is composed of 11 exons from which 10 are transcribed in a 393 amino acids protein. The distribution of the over 25,000 variants reported in the TP53 database available either in a UMD (http://www.umd.be) or LOVD format (http://www.lovd.nl; see paragraph 22.7) shows that approximately 95% of mutations are located in 4 out of the 11 exons of the gene (exons 5−8; Beroud and Soussi, 2003; Soussi and Beroud, 2003). This observation has led 39% of research groups to search for mutations only in these exons, whereas 13% performed a complete scanning of the gene (Soussi and Beroud, 2001). Although this strategy is cost-effective, one needs to be careful in case of a negative result and should perform a complete scanning in order to avoid bias, as described in Soussi and Beroud (2001).

Similarly, summary listings of variations documented in LSDBs (such as those provided in LOVD-based LSDBs) help to choose the best experimental approach. For example, if most variants are nonsense, the protein truncation test could be one of the best approaches, whereas sequencing is considered to be the golden standard for variant detection. Certain LSDBs may contain additional information about primers and technical conditions to help new research groups or diagnostic laboratories establish their own diagnostic procedures, such as those for the various genes involved in muscular dystrophies (http://www.dmd.nl) and thalassemias (http://www.goldenhelix.org/xprbase; a companion database to HbVar; Giardine et al., 2007a).

Finally, in a few LSDBs, explicit information is provided regarding the mutation pattern associated with a population/ethnic group and/or geographical region. This information can be extremely helpful in stratifying mutation detection strategies. In other words, in ethnic groups with a more or less homogeneous mutation pattern, mutation screening efforts can be either targeted to those genomic regions where the majority of mutations have been reported or to a specific mutation detection technique, i.e., amplification-refractory mutation system or restriction enzyme analysis, thus saving time and resources. If available, coupling LSDB with NEMDB resources would further facilitate these efforts. However, as previously explained, extreme caution should be taken in case of a negative result that would require complete mutation scanning in order to avoid bias.

20.5.4 Comparison Among Various Locus-Specific Databases

Several hundred LSDBs are available on the Internet, which sometimes makes it difficult to choose the “best” one. In addition, there is often more than one LSDB per gene locus, from which it is hard to determine the

"reference" and best-curated LSDB. This fact generates confusion to potential users as to which LSDB to choose. This can be particularly worrying, since not all databases conform to the proposed quality guidelines or are curated or updated frequently, and they are rather diverse in terms of content and structure.

In 2010, Mitropoulou et al. performed a thorough domain analysis of the 1,188 existing LSDBs in an effort to comprehensively map data models and ontology options, on which existing LSDBs are based, in order to provide insight into ways the field should further develop and to produce recommendations toward the implementation of LSDBs for use in a clinical and genetic laboratory setting. This effort came as a follow up of the comparative analysis of Claustres et al. (2002), dictated by the rapid growth of LSDBs and the vast data content heterogeneity that characterizes the field. These LSDBs were assessed for a total of 44 content criteria pertaining to general presentation, locus-specific information, database structure, data collection, variant information tables, and database querying.

A key observation that derived from this analysis is the fact that more and more LSDBs are generated using an often downloadable LSDB management system, hence adequately tackling the issue of data—content heterogeneity that existed in the dawn of the LSDB era in the early 2000s. Overall, several elements have helped to advance the field and reduce data heterogeneity, such as the development of specialized DBMS and the creation of improved data querying tools, while a number of deficiencies were identified, namely the lack of detailed disease and phenotypic descriptions for each genetic variant and links to relevant patient organizations, which, if addressed, would allow LSDBs to better serve the clinical genetics community. Based on these findings, Mitropoulou et al. (2010) proposed an LSDB-based structure, which would contribute to a federated genetic variation browser and also allow the seamless maintenance of genomic variation data. To this end, this effort constituted a formal "requirements analysis" that was undertaken by the GEN2PHEN project (http://www.gen2phen.org), aiming to contribute guidelines upon which the LSDB field can be further evolved, formalize the data models and the nomenclature systems being utilized by the entire LSDB community, and maximize synergy with groups involved in the LSDB field.

20.6 NATIONAL/ETHNIC MUTATION DATABASES: ARCHIVING THE GENOMIC BASIS OF HUMAN DISORDERS ON A POPULATION BASIS

NEMDBs are genomic variant repositories, recording extensive information over the described genetic heterogeneity of an ethnic group or population. These resources emerged mostly driven by the need to document the varying mutation spectra observed for any gene (or multiple genes) associated with a genetic disorder among different population and ethnic groups (Patrinos, 2006).

In general, NEMDBs can be divided in two subcategories. The "National Genetic" (or Disease Mutation) databases, the first ones that appeared online, record the extant genetic composition of a population or ethnic group but with limited or no description of mutation frequencies. The first NEMDB to come online was the Finnish database (http://www.findis.org; Sipila and Aula, 2002), which was rich in information but provided a very limited querying capacity, particularly for allelic frequencies. On the other hand, the "National Mutation Frequency" databases provide comprehensive information only of those inherited, mostly monogenic, disorders whose disease-causing mutation spectrum is well defined. The Hellenic and Cypriot NEMDBs (Patrinos et al., 2005a; Kleanthous et al., 2006; respectively) introduced a specialized database management system for NEMDBs that enabled both basic query formulation and restricted-access data entry so that all records are manually curated to ensure high and consistent data quality (van Baal et al., 2010).

In order to provide a simple and expandable system for worldwide population-specific mutation frequency data documentation, the latter NEMDB group was used as the basis for the design of the Frequency of INherited Disorders database (FINDbase; http://www.findbase.org), a relational database that records frequencies of clinically relevant genomic variants worldwide (van Baal et al., 2007). Even from its first version, FINDbase offered a user-friendly query interface, providing instant access to the list and frequencies of the different mutations and query outputs that can be either in a table or graphical format.

In 2010, FINDbase migrated to the new version of the ETHNOS software, which included new data querying and visualization tools to further exploit the expanded FINDbase data collection (Georgitsi et al., 2011a). The data querying and visualization tools were built around Microsoft's PivotViewer software (http://www.getpivot.com), based on Microsoft Silverlight technology (http://www.silverlight.net), which provides an elegant, web-based multimedia interface for population-based genetic variation data collection and retrieval. The whole system architecture is based on a three-tier client-server model (Eckerson, 1995), namely the client application, the application server, and the database server. All of the FINDbase data records were converted to a set of files on a server, which were CXML- and Deep Zoom-formatted images (http://msdn.microsoft.com/en-us/library/cc645077%28VS.95%29.aspx). When the user browses the collection from a webpage, the PivotViewer uses the Silverlight Control to display the files. The entire FINDbase causative mutations data collection via PivotViewer is shown in Fig. 20.1, which enables the user to simultaneously interact with large data sets. PivotViewer enables users to smoothly and quickly arrange FINDbase data collections according to

common characteristics that can be selected from the data query menu (Georgitsi et al., 2011a) and then zoom in for a closer look by either filtering the collection to get a subset of information or clicking on a particular display item. Each display item, in the form of a card (Fig. 20.2) with a chromosomal figure (derived from http://www.genecards.org) displaying the gene position, is provided for each genetic variation, along with a sidebar text box with in-depth data concerning the particular genetic variation and population. Hyperlinks to the OMIM database and HGMD for each gene name offer the user the possibility of easily accessing additional information.

In particular, the new FINDbase data querying and visualization environment enables the user to perform simple and complex queries, visualize and sort, organize and categorize data dynamically, and discover trends across all items, using different views. In 2012, additional visualization tools were implemented (Viennas et al., 2012; Papadopoulos et al., 2014) based on the Flare visualization toolkit (http://flare.prefuse.org), which provides two extra types of data query and visualization outputs, namely the Gene and Mutation Map and the Mutation Dependency Graph. The Gene and Mutation Map is based on a tree-map, which constitutes an easy way of analyzing large data sets (http://www.cs.umd.edu/hcil/treemap). In FINDbase, the tree-map corresponds to mutation frequencies estimated for each population. Each rectangle represents a population's mutation, and a specific color corresponds to each population. The area covering each node encodes the frequency of rare alleles. Each time the user clicks on a node, the occurrence of the selected mutation is shown over all populations. Similarly, the Mutation Dependency Graph visualizes the dependencies that occur among different populations on the basis of a selected genomic variant. In FINDbase, population names are placed along a circle, and these populations are clustered based on the presence and/or the frequency of a certain genomic variant. A link between populations indicates that these populations have the same genomic variant in common, and by clicking on a specific population, the user can see all of the relevant dependencies of that population concerning the selected genomic variant. These tools, available only for the causative mutation FINDbase module, provide the means of establishing relationships among different populations on the basis of certain genome variants and together with the basic data querying option provided by the PivotViewer significantly enhance the battery of data visualization tools available to explore FINDbase data content.

Apart from the documentation of causative genomic variants, leading to inherited disorders, two additional modules have been implemented in FINDbase:

1. *The Pharmacogenomics Biomarker module*, which comprehensively documents the incidence of pharmacogenomics biomarkers in different populations (Georgitsi et al., 2011b). This is a much-needed addition, since population- and ethnic group-specific allele frequencies of pharmacogenomic markers are poorly documented and not systematically collected in structured data repositories.
2. *The Genetic Disease Summaries module*, which is in fact the evolution of the first flat-file ETHNOS-based NEMBDs (van Baal et al., 2010). As with existing

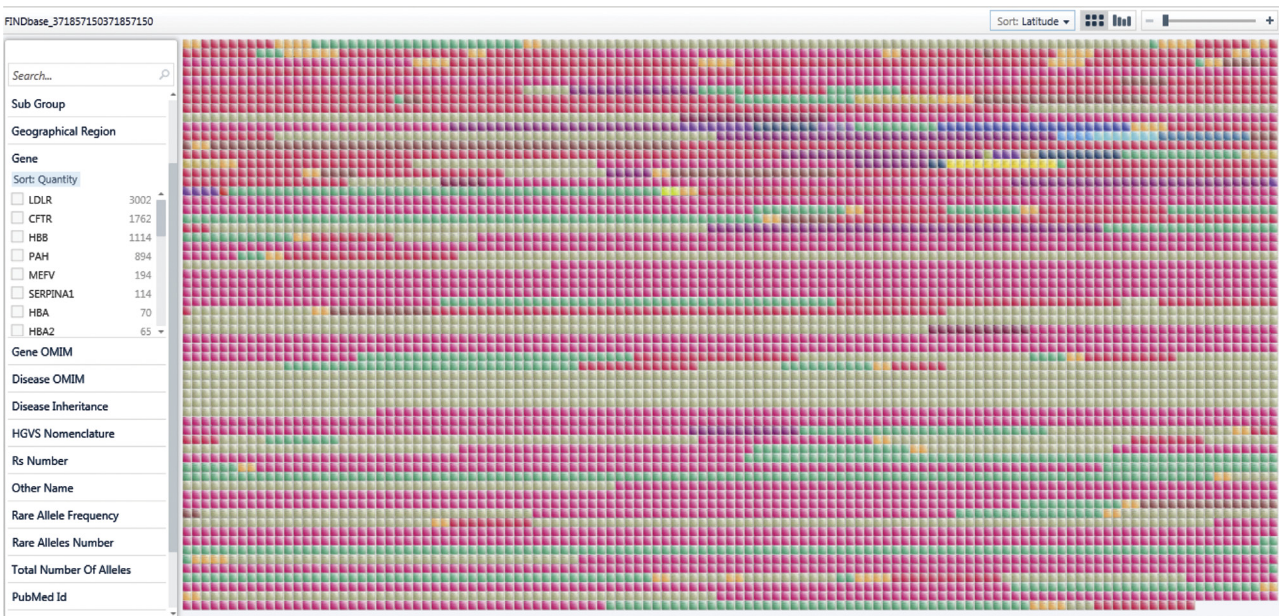


FIGURE 20.1 Overview of the newly designed FINDbase causative mutation module data collection, based on Microsoft's PivotViewer and Silverlight technology. The querying interface is shown on the left, and the output option can be selected at the top right corner of the screen. The different entries are shown as colored boxes, presented as display items (see Fig. 20.2). The user can zoom in for a closer look or click on a particular item to get more in-depth information.

FIGURE 20.2 Example of a display item provided for each causative mutation (*SERPINA1* c.863 A>T), leading to a_1-antitrypsin deficiency, accompanied by a sidebar text box with in-depth data concerning the particular variant and population (Italy; shown on the right). Each item includes the name of the allele in its official Human Genome Variation Society (or other) nomenclature system, if available; the population for which this information is available (shown by the country's flag); and a chromosomal map, where the gene's position is indicated. Hyperlinks for each gene to the Online Mendelian Inheritance in Man database offer the user the possibility of easily accessing additional information. Finally, each item displays the corresponding PubMed and Researcher IDs, if applicable. Similar display items are also used for the Pharmacogenomic Biomarkers and Genetic Disease Summaries modules (not shown).

FINDbase modules for causative mutations and pharmacogenomic biomarkers, the component services of this module were also built using the same principles and exploiting PivotViewer and Microsoft Silverlight technology. In this module, database records include the population, the genetic disease, the gene name, the OMIM ID, and the chromosome on which it resides. These are all included in the query interface.

To maximize the utility of NEMDBs, given their scope, the mode by which their content is provided needs to provide a seamless integration with related content in LSDBs and GMDs (Patrinos and Brookes, 2005). Furthermore, as is always desirable for specialized databases, extensive links to other external information (e.g., to OMIM and to various types of genome sequence annotations) would ideally be provided as part of the necessary tying together of the growing network of genomic databases.

20.6.1 National/Ethnic Mutation Databases and Genetic Testing

NEMDBs can be helpful in multiple ways in a molecular diagnostic setting. First of all, they can help optimize national molecular diagnostic services by providing essential reference information for the design and implementation of regional or national mutation screening efforts (see also Section 20.4). In addition, NEMDBs can enhance awareness among clinicians, bioscientists, and the general public about the range of the most common genetic disorders suffered by certain populations and/or ethnic groups. Based on populations' mutation spectra, customized genetic tests can be designed. Several diagnostic companies have designed kits for β-thalassemia, which qualitatively detect the most common Mediterranean and Asian β-thalassemia mutations (reviewed in Patrinos et al., 2005b). Most importantly, these databases can also assist in interpreting diagnostic test results in countries with heterogeneous

populations, particularly where the interpretation of test results in minority ethnic groups may be ambiguous or problematic (Zlotogora et al., 2007, 2009).

20.6.2 National/Ethnic Mutation Databases and Society

Apart from their importance in a clinical laboratory setting, NEMDBs can also contribute toward the elucidation of populations' origins and migrations. The history of a certain population is tightly linked with the history of its allele(s). Therefore NEMDBs, particularly those including data from many population groups, can serve as the platform for comparative genomic studies that can reciprocally provide insights into, for example, the demographic history of human populations, patterns of their migration and admixture, gene/mutation flow, etc. NEMDBs can therefore be particularly helpful in determining the microevolution of human populations via precise ethnic identification that in turn is extremely important for the rapid transition of modern societies, through migration, into multiethnic societies.

To this end, in order for an NEMDB to be maximally comprehensive and accurate, it is vital that certain parameters are strictly observed. In the case of recurrent mutation events (e.g., sickle cell mutation), caution should be taken to precisely record the underlying genomic background on which a mutation has occurred. Furthermore, mutation frequencies should be calculated based on the most representative study that involves sufficient numbers of patients and controls. The estimation of absolute mutation frequencies based on multiple reports has the inherent danger of including redundant cases that can alter the calculated allele frequencies. Finally, the very delicate issue of anonymity should be adequately preserved by including data only at the summary rather than the individual level, so that NEMDB data contents consist only of a number of chromosomes rather than sensitive personal details of their carriers. Several recommendations and guidelines to facilitate the participation of emerging countries in genetic variation data documentation, ensuring an accurate and comprehensive worldwide data collection, have been proposed (Patrinos et al., 2011): (1) developing procedures to include clinicians and researchers in developing nations; (2) creating international networks from closely related populations to enhance communication and interaction and initiate research networks between them; (3) fostering communication, interaction, and research networks between developing and developed countries; (4) ensuring that the DBMS that are being used or developed can be utilized by those in a limited resource environment; (5) providing support to developing countries to build capacity and to fully participate in the collection, analysis, and sharing of genetic variation information; (6) developing a framework to facilitate interactions between the coordinating center and national, regional, and international agencies; and (7) ensuring that all ethical, legal, religious, and social issues are thoroughly considered when NEMDBs and/or data capture projects are launched in developing countries. Based on these recommendations, a number of Country Nodes have been developed for the Human Variome Project (HVP; Patrinos et al., 2012b).

20.7 DATABASE MANAGEMENT SYSTEMS FOR LOCUS-SPECIFIC DATABASES AND NATIONAL/ETHNIC MUTATION DATABASES

Ever since the development of the first genetic databases, a number of guidelines and recommendations for mutation nomenclature (den Dunnen and Antonarakis, 2001; den Dunnen and Paalman, 2003), content, structure (Scriver et al., 1999), curation, and the deployment of LSDBs (Claustres et al., 2002; Cotton et al., 2008; Mitropoulou et al., 2010) and NEMDBs (Patrinos, 2006, 2012b) have been produced to encourage the harmonization of LSDB and NEMDB development and curation.

To facilitate interested parties and research groups to develop and curate their own LSDBs and NEMDBs, generic tools, known as DBMS, have been made available. Several off-the-shelf, freely available, and user-friendly software packages have been developed for LSDB development and curation, including MUTbase (Riikonen and Vihinen, 1999); UMD (Beroud et al., 2000); Mutation Storage and Retrieval (Brown and McKie, 2000); and LOVD (Fokkema et al., 2005). From these, LOVD, now in V3.0 with a total of 13,125,955 (2,406,719 unique) variants in 83 LOVD installations (last accessed March 2016), in synergy with the HVP (Kaput et al., 2009; Cotton et al., 2009; http://www.humanvariomeproject.org), the National Center for Biotechnology Information (NCBI), and other online resources, is considered by far to be the most commonly used DBMS for LSDB development and curation. This software provides, mostly in a tabular format and in an easy-to-query manner, detailed lists of all variants documented for a specific gene locus (Fig. 20.3), while a summary page allows the user to scan through graphical summaries of all genomic variants available in a specific installation (Fig. 20.4). In its newest release, LOVD extends this concept to provide patient-centered data storage and storage of next generation sequencing data, hence allowing for the deposition of variants that reside in intergenic regions with possible functional significance.

As far as NEMDBs are concerned, ETHNOS (van Baal et al., 2010) is available, on which the first-generation NEMDBs and FINDbase were developed. Table 20.1 summarizes the existing DBMS available today.

LOVD Gene homepage

General information	
Gene name	Kruppel-like factor 1 (erythroid)
Gene symbol	**KLF1**
Chromosome Location	19p13.13-p13.12
Database location	lovd.bx.psu.edu
Curator	Belinda Giardine and Joseph Borg
PubMed references	View all (unique) PubMed references in the KLF1 database
Date of creation	May 06, 2010
Last update	November 12, 2014
Version	**KLF1 141112**
Add sequence variant	Submit a sequence variant
First time submitters	Register here
Genomic refseq ID	NC_000019.9
Transcript refseq ID	NM_006563.3
Total number of unique DNA variants reported	**36**
Total number of individuals with variant(s)	**30**
Total number of variants reported	**39**
Subscribe to updates of this gene	

Graphical displays and utilities	
Summary tables	Summary of all sequence variants in the KLF1 database, sorted by type of variant (with graphical displays and statistics)
UCSC Genome Browser	Show variants in the UCSC Genome Browser (compact view)
PSU Genome Browser	Show variants in the PSU Genome Browser (compact view)
Ensembl Genome Browser	Show variants in the Ensembl Genome Browser (compact view)
NCBI Sequence Viewer	Show distribution histogram of variants in the NCBI Sequence Viewer

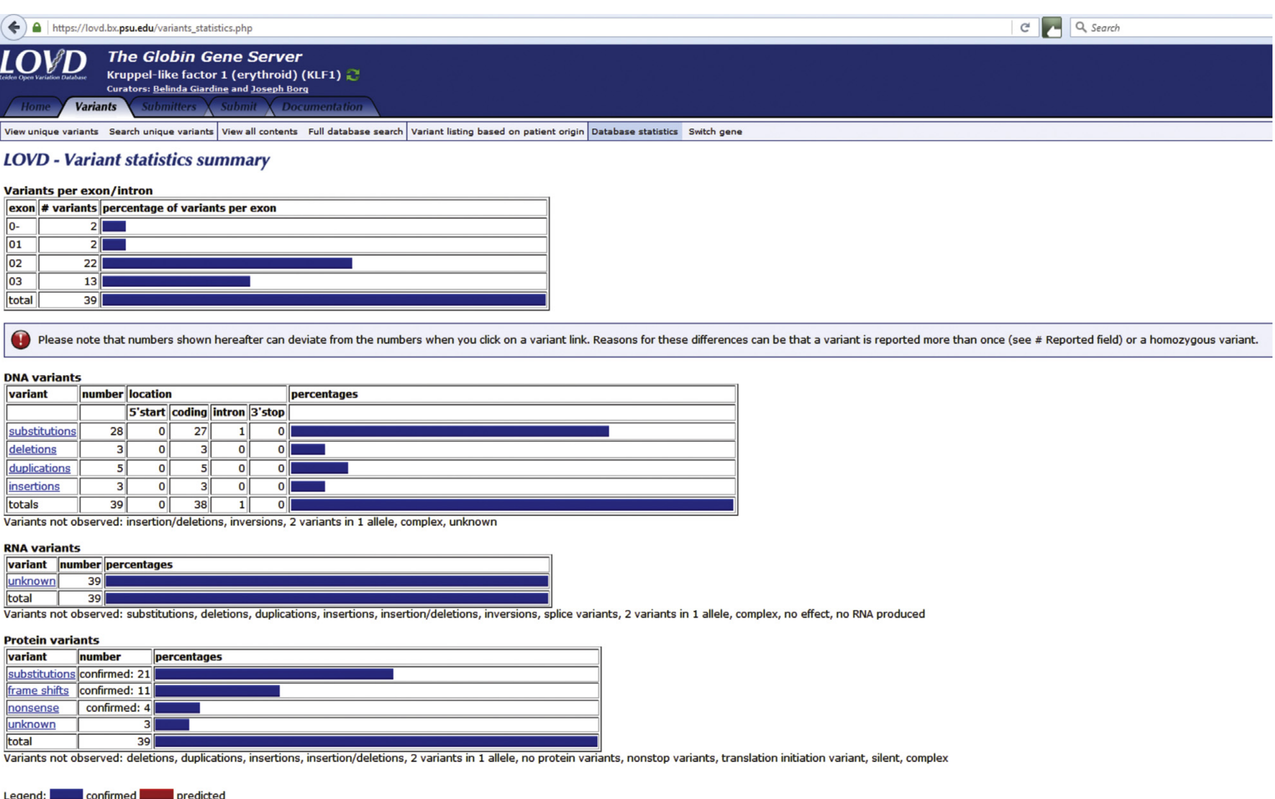
Sequence variant tables	
Unique sequence variants	Listing of all unique sequence variants in the KLF1 database, without patient data
Complete sequence variant listing	Listing of all sequence variants in the KLF1 database
Variants with no known pathogenicity	Listing of all KLF1 variants reported to have no noticeable phenotypic effect (note: excluding variants of unknown effect)
Download table	Download the full sequence variant table of the KLF1 database in tab-delimited text format.

Search the database	
By type of variant	View all sequence variants of a certain type
Simple search	Query the database by selecting the most important variables (exon number, type of variant, disease phenotype)
Advanced search	Query the database by selecting a combination of variables
Based on patient origin	View all variants based on your patient origin search terms

Links to other resources	
Entrez Gene	10661
OMIM - Gene	600599
UniProtKB (SwissProt/TrEMBL)	Q13351
External link	HbVar

FIGURE 20.3 Overview of a typical Leiden-Open (source) Variation Database gene homepage (in this example for the *KLF1* gene). Detailed information regarding the total number of variants documented and the names and contact details of the database curators are provided along with hyperlinks to related resources, such as the Online Mendelian Inheritance in Man database, NCBI, etc. Graphical displays are also provided in conjunction with other resources.

These user-friendly DBMS are designed to promote the creation of more and better LSDBs by reducing or eliminating the requirement of substantial knowledge of computing and bioinformatics for interested parties to establish an LSDB from scratch. In addition, the use of off-the-shelf solutions positively impacts data uniformity and, contrary to the NEMDBs structure and data content that is relatively uniform since this discipline has more recently


https://lovd.bx.psu.edu/variants_statistics.php

LOVD Leiden Open Variation Database

The Globin Gene Server

Kruppel-like factor 1 (erythroid) (KLF1)

Curators: Belinda Giardine and Joseph Borg

Home | Variants | Submitters | Submit | Documentation

View unique variants | Search unique variants | View all contents | Full database search | Variant listing based on patient origin | Database statistics | Switch gene

LOVD - Variant statistics summary

Variants per exon/intron

exon	# variants	percentage of variants per exon
0-	2	
01	2	
02	22	
03	13	
total	39	

Please note that numbers shown hereafter can deviate from the numbers when you click on a variant link. Reasons for these differences can be that a variant is reported more than once (see # Reported field) or a homozygous variant.

DNA variants

variant	number	location				percentages
		5'start	coding	intron	3'stop	
substitutions	28	0	27	1	0	
deletions	3	0	3	0	0	
duplications	5	0	5	0	0	
insertions	3	0	3	0	0	
totals	39	0	38	1	0	

Variants not observed: insertion/deletions, inversions, 2 variants in 1 allele, complex, unknown

RNA variants

variant	number	percentages
unknown	39	
total	39	

Variants not observed: substitutions, deletions, duplications, insertions, insertion/deletions, inversions, splice variants, 2 variants in 1 allele, complex, no effect, no RNA produced

Protein variants

variant	number	percentages
substitutions	confirmed: 21	
frame shifts	confirmed: 11	
nonsense	confirmed: 4	
unknown	3	
total	39	

Variants not observed: deletions, duplications, insertions, insertion/deletions, 2 variants in 1 allele, no protein variants, nonstop variants, translation initiation variant, silent, complex

Legend: confirmed predicted

FIGURE 20.4 Screenshot of the variant statistics summary page of a typical Leiden-Open (source) Variation Database webpage (*KLF1* in this case; see also Fig. 20.3). The different types of variants for DNA, RNA, and protein are provided in blue (observed variants) or red (predicted variants). Variants are also clustered per exons and introns, hence allowing for the identification of mutational hotspots, facilitating the design of molecular diagnostics assays.

emerged, the use of such DBMS in the LSDB domain resulted in a significant increase of LSDB structure and data content homogeneity (Mitropoulou et al., 2010), hence drastically solving the issue and increasing data uniformity among these resources.

Finally, the use of DBMS that can be run on any platform will reduce the risk of the database being "lost." If database curation for some reason, e.g., a lack of funding, is interrupted, data will then be transferred directly between platforms or locations, and they will remain accessible to all. Potential curators will be encouraged to set up LSDBs, with the choice of using these software packages locally on their own workstations or, most importantly, having their databases hosted on a central server on the Internet.

20.8 INCENTIVIZING DATA SHARING: THE MICROATTRIBUTION APPROACH

The advent of next-generation sequencing (see Chapter 9) poses new challenges as to which extended individual genomes differ from each other and how phenotypic variation can be attributed to these differences. To address this, DNA sequence data will have to be matched with well-defined phenotypes to make meaningful connections between structure, function, and mechanism. A major hurdle to address this issue is to motivate data submitters to share their findings, since it has become increasingly difficult to report small numbers of human variants in scientific journals.

Traditionally, authors of scientific articles describing genomic variants were rewarded by citations of their publications describing these variants. However, the increased rate of discovery through resequencing efforts far exceeds the capacity of citations of individual publications to give adequate credit. On the other hand, discovered variants can be deposited into genetic databases, such as those described above, but data submitters do not get any credit for sharing their data. To overcome this problem, Giardine et al. (2011) proposed and implemented a process for capturing such information with the incentive of microattribution, whereby the contribution of those collecting new detailed genotype/phenotype data is positively encouraged and appropriately acknowledged and rewarded. As such, microattribution can give credit to the discoverers of variants deposited in databases. At first, all genetic variation data have to be collected and documented in a genetic database with the appropriate attribution of data contributors. All genomic variants should then be deposited into stable, publicly available, and well-maintained central repositories [e.g., NCBI, European

Bioinformatics Institute (EBI)] that would run independent microattribution services based on an individual researcher's unique identity, such as is aimed for by the Open Researcher and Contributor ID consortium (http://orcid.org), ResearcherID (http://www.researcherid.com), OpenID (http://openid.net), etc.

Depositing genome data as "nanopublications" (Mons and Velterop, 2009), contributed by an individual researcher or research group, in a stable and accessible format in open repositories, would allow for the mining for citations associated with this/these authors' unique IDs. In other words, a nanopublication is the smallest unit of publishable information that can be linked to its contributor via his or her unique scientific identity and which can be cited and evaluated in terms of its impact upon the research community. Subsequently, a Microattribution Analysis article, i.e., a comprehensive high-profile article commissioned by the journal's editor and stringently peer-reviewed, would summarize the features of all variants at a particular locus, such as phenotypes, clinical findings, allele frequencies, and so on. In this article, all genome variant contributors who are authors of the nanopublication collection constituting the Microattribution Analysis article would be considered as coauthors or would in the future receive citations on their individual nanopublications.

There are several advantages to this approach. First of all, the nature of the nanopublication is such that it does not need to be related to a full scientific article (although this is also an option); it can refer to a database from which it is derived. Moreover, nanopublications can be encoded in the Resource Description Framework to become part of the Semantic Web and transferred between computers using Extensible Markup Language. As such, nanopublications can be more easily queried and retrieved through the Web, something that is not feasible for regular articles. Most importantly, nanopublications have the potential to incentivize potential data contributors to place their data in the public domain for others to freely access and optimally exploit, since nanopublications can be attributed and cited in the same way as regular articles (Mons et al., 2011). This can have a significant impact upon future scientific publication modalities, since it facilitates data mining, enables data sharing, and increases the likelihood of an article being retrieved and cited.

Microattribution was first implemented using the human globin genes, and the resulting hemoglobinopathies and thalassemias are an example of the systematic documentation of genetic variation leading to human genetic disorders (Giardine et al., 2011). Through this exercise, a multicenter consortium has demonstrated that not only can microattribution incentivize data contribution, but most importantly, it shows how an integrated human variant database can provide key insights into human genetic diseases, which otherwise would not be possible. In other words, Giardine et al. (2011) showed that microattribution provides an important mechanism and incentive for researchers to report all variants within a specific gene or disease network. Following the principles established for the hemoglobinopathies, these databases should provide a key resource for understanding the molecular pathology of human genetic diseases. At the same time, microattribution was successfully implemented to incentivize data sharing in NEMDBs (Georgitsi et al., 2011b) and clinical genetics databases (Sosnay et al., 2013; Thomson et al., 2014a).

Still, there are several issues that should be addressed before microattribution and/or similar initiatives can be successfully implemented in practice (Patrinos et al., 2012b). First, the single most important contribution of microattribution to increase the rate of data submission is to incentivize individual researchers or research groups to submit their newly acquired and unpublished mutation/variation data to a public repository or database in return for appropriate credit and attribution. As such, these initiatives presuppose the existence of well-curated data repositories and underlying mutation data sets as skeleton structures. Second, the successful adoption of microattribution would require its own quality evaluation/peer-review process by international experts to review and curate the submitted data where necessary. In the context of human genome variation data, a sensible approach would be to base the whole process around one or more preexisting, freely available, high-quality centralized databases, as previously proposed (Giardine et al., 2011; Georgitsi et al., 2011b). In relation to this approach, serious consideration must be given to the scale of funding required to run such a venture. One possible funding mechanism would be an article processing charges-type of fee, similar to those imposed in open-access journals, a solution that may easily be provided by partnering a centralized data repository with a major publishing group(s) along the lines of a database-journal-like model (Patrinos and Wajcman, 2004; Patrinos and Petricoin, 2009).

Another challenge in capturing all ascertained genome variation data lies in the incentivization of public and private genetic diagnostic laboratories, provided of course that certain ethical issues are overcome, such as obtaining the necessary consent or establishing a specific policy to conceal one's personal details upon data submission. Certain DBMS have made provisions to contribute patients' genomic data without revealing any sensitive personal details, both in the context of individuals (Fokkema et al., 2011) and of culturally sensitive population groups (Zlotogora et al., 2009). Another practical step that could be implemented to tackle this bottleneck would be to include data contributions to public repositories as a necessary quality control requirement for the accreditation process of a genetic laboratory. Such a measure is already being

TABLE 20.1 Database Management Systems for the Development and Curation of Locus-Specific Databases and National/Ethnic Mutation Databases

DBMS	URL	Installations		References
		2008	2016	
LSDBs				
MUTbase	http://bioinf.uta.fi/MUTbase	N.A.	N.A.	Riikonen and Vihinen (1999)
UMD	http://www.umd.be	18	34	Beroud et al. (2000)
MuStaR	http://www.hgu.mrc.ac.uk/Softdata/Mustar	4	N.A.	Brown and McKie (2000)
LOVD	http://www.lovd.nl	20	86	Fokkema et al. (2011)
NEMDBs				
ETHNOS	http://www.findbase.org	12	90	Van Baal et al. (2010)

N.A., *not available.*

implemented by the Israeli Ministry of Health (Zlotogora et al., 2009).

Microattribution should also be implemented in such a way as to inspire busy, high-profile researchers and their groups to populate LSDBs with genome variation data. In other words, it may be possible to not only track a researcher's contribution as a means of providing publication credit but also establish a microcredit measure for database development, maintenance, and curation efforts, involving, for example, manual data content curation and expansion by adding related information from the literature, verifying data correctness, and cross-linking/partnering with other databases. At present, such efforts go largely unnoticed, since they cannot be measured by traditional publication metrics. This requires new mechanisms for citation metrics that will be adopted on top of already existing ones (e.g., H-index, etc.). As such, the technical ability to cite and attribute small yet significant contributions to -omics information in general is necessary but not sufficient to change the culture of attribution and reward in the biological sciences.

Even though the value of incremental micro-contributions to the scientific knowledge base is widely recognized by the research community itself, the reward systems used by funders and faculty bodies to decide upon issues of tenure and promotion are not yet ready to accommodation this shift. It is therefore imperative that, alongside technical changes, social changes are made to make the scientific reward system ready for the next decades, including taking into account the crucial role that curated resources such as LSDBs, NEMDBs, and central databases play in modern genomics and translational research and the vital need for fundraising to keep these resources well maintained and sustainable for the long term (and hence viable). This latter requirement is infinitely more significant, since an objective database quality measure for reviewers to assess is the number of visits and overall citations to the reports describing the database itself or the database traffic metrics.

20.9 FUTURE CHALLENGES

Notwithstanding the technical challenges, perhaps more difficult to overcome will be problems associated with the way not only data submission is rewarded (see previous paragraph) but also how database research is organized and motivated. For example, forming consensus opinions and truly committed consortia in order to create standards is far from easy in the highly competitive world of science. This may partly explain why leading bioinformatics activities are often conducted in large, specialized centers (e.g., the EBI and the US NCBI) where the political influence and critical mass is such that what they produce automatically becomes the de facto standard. These groups, however, cannot build all of the necessary GMDs, LSDBs, and NEMDBs that are needed. They could, however, help others (biological domain experts) to do it and then integrate all of their efforts (Stein, 2002).

In 2006, the HVP was globally initiated, aiming to catalog all human genetic variations and to make that information freely available to researchers, clinicians, and patients worldwide (Ring et al., 2006). The HVP (http://www.humanvariomeproject.org) envisions achieving improved health outcomes by facilitating the unification of human genetic variation and its impact on human health (Horaitis et al., 2007). It will support the use of human variation information in clinical and research environments across the world by developing the resources required to undertake the following key tasks: (1) to capture and archive all human gene variations associated with human diseases developed from gene-specific curation in a central location with mirror sites in other

countries to maximize data security and integrity, allowing for searching across all genes using a common interface; (2) to establish systems that ensure the adequate curation of human variation knowledge from LSDBs, NEMDBs, or disease-specific database perspectives to improve accuracy, reduce errors, and develop a comprehensive data set comprising all human genes; (3) to facilitate the development of software to collect and exchange human variation data in a federation of GMDs, LSDBs, NEMDBs, and disease-specific databases; (4) to create a support system for research laboratories for the collection of genotypic and phenotypic data together using the defined reference sequence in a free, unrestricted, and open access system and create a simple mechanism for logging discoveries; (5) to develop ethical standards that ensure open access to all human variation data that are to be used for global public good and address the needs of indigenous communities under threat of dilution in emerging countries; (6) to provide support to developing countries to build capacity and to fully participate in the collection, analysis, and sharing of genetic variation information; and (7) to establish a communication and education program to collect and spread knowledge related to human variation knowledge to all countries of the world (Kaput et al., 2009; Cotton et al., 2009).

An enormous amount of content already exists in the digital universe, characterized by high rates of new information that is added, distributed, and demands attention. This big data momentum lacks applicability and societal impact at significant levels, as there are several gaps at various knowledge levels that set individuals and experts apart. Indicatively, genetic information cannot be equally understood by a researcher, a geneticist, or a clinician. For this, IT tools are anticipated to play a crucial role toward data management and decision making. In this context, DruGeVar (Dalabira et al., 2014) was developed as an online resource to triangulate drugs with genes and variants that serve as pharmacogenomic biomarkers. On the basis of such a clinical pharmacogenomics initiative, an electronic pharmacogenomics assistant would be of value. Potamias and coworkers developed such a system to provide personalized drug recommendations based on the genotype-to-phenotype data presented and validated by regulatory agencies. Of course, such an electronic pharmacogenomics assistant may assist biomedical research toward the identification of new gene-variant-biomarker entities (Potamias et al., 2014).

Finally, the most fundamental hurdle of all that retards the field is that of limited funding. Because of this, almost all mutation databases in existence have been built by researchers "on the side" for their own use, with a small degree of sponsorship/funding at best. To advance beyond this cottage industry state of affairs, projects need to be increased in scale, quality, and durability, and this can only happen if strategically minded funding agencies make available substantial targeted funds. The new databases that emerge will then need long-term support for general maintenance and further development. To solve this, the projects may ultimately need to be run as self-sustaining "businesses" that charge for data access, and/or it might be possible to develop novel forms of joint academic-corporate funding.

With great vision, in 2007 the European Commission announced the first call for proposals for the seventh Framework Program (FP7; thematic area HEALTH), which included a topic on "…unifying human and model organism genetic variation databases." This topic focused on developing a data and analysis structure by creating a hierarchy of bioinformatics grid-linked databases, tools, and standards, centered on a generalized existing or novel genome browser. GEN2PHEN (2007−13; http://www.gen2phen.org), a large-scale project with 19 participating academic and corporate entities, was funded from this call for proposals, aiming to unify human genetic variation databases toward increasingly holistic views into genotype-to-phenotype (G2P) data and to link this system to other biomedical knowledge sources via genome browser functionality. The GEN2PHEN strategy was designed to do the following: (1) analyze the G2P field and investigate current needs and practices; (2) develop key standards for the G2P field; (3) create generic database components, services, and integration infrastructures for the G2P domain; (4) create data search and presentation solutions for G2P knowledge; (5) facilitate the populating of research and diagnostic G2P databases; (6) build a major G2P internet portal; (7) deploy GEN2PHEN solutions to the community; and (8) address system durability and long-term financing. Consequently, a similar FP7 project (RD-Connect; http://www.rd-connect.eu; Thomson et al., 2014b) is envisaged to act as a warehouse for bioinformatics solutions to facilitate research and discovery in rare diseases. Similar funding opportunities have also been announced from other funding bodies, such as the National Institutes of Health in the United States, aiming to address this need.

20.10 CONCLUSIONS

It is widely accepted that LSDBs and NEMDBs are increasingly becoming valuable tools in molecular diagnostics. Although there has been a significant improvement in both qualitative (data uniformity and database quality) and quantitative terms (an increase in the number of LSDBs and NEMDBs), there are still limitations in the degree of interconnection of these resources to capture all that is known and being discovered regarding pathogenic DNA variants. The main reason for this deficiency is that the modern research ethos fails to provide adequate incentives (i.e., publication options, peer recognition, funding) to encourage researchers to build new and/or curate existing databases. Apparently, the

biomedical community must first appreciate the overwhelming need for further improving genetic/mutation database systems, and the most adequate solution will then presumably follow.

ACKNOWLEDGMENTS

Most of our work has been supported by funds from the European Commission grants (GEN2PHEN FP7-200754, RD-Connect FP7-305444) to George P. Patrinos.

REFERENCES

Agarwala, R., Barrett, T., Beck, J., Benson, D.A., Bollin, C., Bolton, E., Bourexis, D., Brister, J.R., Bryant, S.H., Canese, K., Charowhas, C., Clark, K., DiCuccio, M., Dondoshansky, I., Federhen, S., Feolo, M., Funk, K., Geer, L.Y., Gorelenkov, V., Hoeppner, M., Holmes, B., Johnson, M., Khotomlianski, V., Kimchi, A., Kimelman, M., Kitts, P., Klimke, W., Krasnov, S., Kuznetsov, A., Landrum, M.J., Landsman, D., Lee, J.M., Lipman, D.J., Lu, Z., Madden, T.L., Madej, T., Marchler-Bauer, A., Karsch-Mizrachi, I., Murphy, T., Orris, R., Ostell, J., O'Sullivan, C., Panchenko, A., Phan, L., Preuss, D., Pruitt, K.D., Rodarmer, K., Rubinstein, W., Sayers, E.W., Schneider, V., Schuler, G.D., Sherry, S.T., Sirotkin, K., Siyan, K., Slotta, D., Soboleva, A., Soussov, V., Starchenko, G., Tatusova, T.A., Todorov, K., Trawick, B.W., Vakatov, D., Wang, Y., Ward, M., Wilbur, W.J., Yaschenko, E., Zbicz, K., 2016. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 44, D7–D19.

Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F., Hamosh, A., 2015. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic Acids Res. 43 (Database issue), D789–D798.

Beroud, C., Soussi, T., 2003. The UMD-p53 database: new mutations and analysis tools. Hum. Mutat. 21, 176–181.

Beroud, C., Collod-Beroud, G., Boileau, C., Soussi, T., Junien, C., 2000. UMD (Universal mutation database): a generic software to build and analyze locus-specific databases. Hum. Mutat. 15, 86–94.

Brown, A.F., McKie, M.A., 2000. MuStaR and other software for locus-specific mutation databases. Hum. Mutat. 15, 76–85.

Claustres, M., Horaitis, O., Vanevski, M., Cotton, R.G., 2002. Time for a unified system of mutation description and reporting: a review of locus-specific mutation databases. Genome Res. 12, 680–688.

Collod, G., Beroud, C., Soussi, T., Junien, C., Boileau, C., 1996. Software and database for the analysis of mutations in the human FBN1 gene. Nucleic Acids Res. 24, 137–140.

Collod-Beroud, G., Boileau, C., 2002. Marfan syndrome in the third Millennium. Eur. J. Hum. Genet. 10, 673–681.

Collod-Beroud, G., Le Bourdelles, S., Ades, L., Ala-Kokko, L., Booms, P., Boxer, M., Child, A., Comeglio, P., De Paepe, A., Hyland, J.C., Holman, K., Kaitila, I., Loeys, B., Matyas, G., Nuytinck, L., Peltonen, L., Rantamaki, T., Robinson, P., Steinmann, B., Junien, C., Beroud, C., Boileau, C., 2003. Update of the UMD-*FBN1* mutation database and creation of an *FBN1* polymorphism database. Hum. Mutat. 22, 199–208.

Cooper, D.N., Ball, E.V., Krawczak, M., 1998. The human gene mutation database. Nucleic Acids Res. 26, 285–287.

Cotton, R.G., McKusick, V., Scriver, C.R., 1998. The HUGO mutation database initiative. Science 279, 10–11.

Cotton, R.G., Phillips, K., Horaitis, O., 2007. A survey of locus-specific database curation. Human Genome Variation Society. J. Med. Genet. 44, e72.

Cotton, R.G., Auerbach, A.D., Beckmann, J.S., Blumenfeld, O.O., Brookes, A.J., Brown, A.F., Carrera, P., Cox, D.W., Gottlieb, B., Greenblatt, M.S., Hilbert, P., Lehvaslaiho, H., Liang, P., Marsh, S., Nebert, D.W., Povey, S., Rossetti, S., Scriver, C.R., Summar, M., Tolan, D.R., Verma, I.C., Vihinen, M., den Dunnen, J.T., 2008. Recommendations for locus-specific databases and their curation. Hum. Mutat. 29, 2–5.

Cotton, R.G., Al Aqeel, A.I., Al-Mulla, F., Carrera, P., Claustres, M., Ekong, R., Hyland, V.J., Marafie, M.J., Paalman, M.H., Patrinos, G.P., Qi, M., Ramesar, R.S., Scott, R.J., Sijmons, R.H., Sobrido, M.J., Vihinen, M., 2009. Capturing all disease-causing mutation for clinical and research use: towards an effortless system for the Human Variome Project. Genet. Med. 11, 843–849.

Dalabira, E., Viennas, E., Daki, E., Komianou, A., Bartsakoulia, M., Poulas, K., Katsila, T., Tzimas, G., Patrinos, G.P., 2014. DruGeVar: an online resource triangulating drugs with genes and genomic biomarkers for clinical pharmacogenomics. Public Health Genom. 17 (5–6), 265–271.

den Dunnen, J.T., Antonarakis, S.E., 2001. Nomenclature for the description of human sequence variations. Hum. Genet. 109, 121–124.

den Dunnen, J.T., Paalman, M.H., 2003. Standardizing mutation nomenclature: why bother? Hum. Mutat. 22, 181–182.

Eckerson, W.W., 1995. Three tier client/server architecture: achieving scalability, performance, and efficiency in client server applications. Open Inf. Syst. 10, 3(20).

Fokkema, I.F., den Dunnen, J.T., Taschner, P.E., 2005. LOVD: easy creation of a locus-specific sequence variation database using an "LSDB-in-a-box" approach. Hum. Mutat. 26, 63–68.

Fokkema, I.F., Taschner, P.E., Schaafsma, G.C., Celli, J., Laros, J.F., den Dunnen, J.T., 2011. LOVD v.2.0: the next generation in gene variant databases. Hum. Mutat. 32, 557–563.

George, R.A., Smith, T.D., Callaghan, S., Hardman, L., Pierides, C., Horaitis, O., Wouters, M.A., Cotton, R.G., 2008. General mutation databases: analysis and review. J. Med. Genet. 45, 65–70.

Georgitsi, M., Viennas, E., Gkantouna, V., van Baal, S., Petricoin, E.F., Poulas, K., Tzimas, G., Patrinos, G.P., 2011a. FINDbase: a worldwide database for genetic variation allele frequencies updated. Nucleic Acids Res. 39, D926–D932.

Georgitsi, M., Viennas, E., Gkantouna, V., Christodoulopoulou, E., Zagoriti, Z., Tafrali, C., Ntellos, F., Giannakopoulou, O., Boulakou, A., Vlahopoulou, P., Kyriacou, E., Tsaknakis, J., Tsakalidis, A., Poulas, K., Tzimas, G., Patrinos, G.P., 2011b. Population-specific documentation of pharmacogenomic markers and their allelic frequencies in FINDbase. Pharmacogenomics 12, 49–58.

Gerlai, R., 2002. Phenomics: fiction or the future? Trends Neurosci. 25, 506–509.

Giardine, B., Riemer, C., Hefferon, T., Thomas, D., Hsu, F., Zielenski, J., Sang, Y., Elnitski, L., Cutting, G., Trumbower, H., Kern, A., Kuhn, R., Patrinos, G.P., Hughes, J., Higgs, D., Chui, D., Scriver, C., Phommarinh, M., Patnaik, S.K., Blumenfeld, O., Gottlieb, B., Vihinen, M., Väliaho, J., Kent, J., Miller, W., Hardison, R.C., 2007a. PhenCode: connecting ENCODE data with mutations and phenotype. Hum. Mutat. 28, 554–562.

Giardine, B., van Baal, S., Kaimakis, P., Riemer, C., Miller, W., Samara, M., Kollia, P., Anagnou, N.P., Chui, D.H., Wajcman, H., Hardison, R.C., Patrinos, G.P., 2007b. HbVar database of human

hemoglobin variants and thalassemia mutations: 2007 update. Hum. Mutat. 28, 206.

Giardine, B., Borg, J., Higgs, D.R., Peterson, K.R., Philipsen, S., Maglott, D., Singleton, B.K., Anstee, D.J., Basak, A.N., Clark, B., Costa, F.C., Faustino, P., Fedosyuk, H., Felice, A.E., Francina, A., Galanello, R., Gallivan, M.V., Georgitsi, M., Gibbons, R.J., Giordano, P.C., Harteveld, C.L., Hoyer, J.D., Jarvis, M., Joly, P., Kanavakis, E., Kollia, P., Menzel, S., Miller, W., Moradkhani, K., Old, J., Papachatzopoulou, A., Papadakis, M.N., Papadopoulos, P., Pavlovic, S., Perseu, L., Radmilovic, M., Riemer, C., Satta, S., Schrijver, I., Stojiljkovic, M., Thein, S.L., Traeger-Synodinos, J., Tully, R., Wada, T., Waye, J.S., Wiemann, C., Zukic, B., Chui, D.H., Wajcman, H., Hardison, R.C., Patrinos, G.P., 2011. Systematic documentation and analysis of human genetic variation in hemoglobinopathies using the microattribution approach. Nat. Genet. 43, 295−301.

Giardine, B., Borg, J., Viennas, E., Pavlidis, C., Moradkhani, K., Joly, P., Bartsakoulia, M., Riemer, C., Miller, W., Tzimas, G., Wajcman, H., Hardison, R.C., Patrinos, G.P., 2014. Updates of the HbVar database of human hemoglobin variants and thalassemia mutations. Nucleic Acids Res. 42, D1063−D1069.

Hall, J.G., 2003. A clinician's plea. Nat. Genet. 33, 440−442.

Hardison, R.C., Chui, D.H., Giardine, B., Riemer, C., Patrinos, G.P., Anagnou, N., Miller, W., Wajcman, H., 2002. HbVar: a relational database of human hemoglobin variants and thalassemia mutations at the globin gene server. Hum. Mutat. 19, 225−233.

Horaitis, O., Talbot Jr., C.C., Phommarinh, M., Phillips, K.M., Cotton, R.G., 2007. A database of locus-specific databases. Nat. Genet. 39, 425.

International HapMap Consortium, 2003. The International HapMap Project. Nature 426, 789−796.

Kaput, J., Cotton, R.G., Hardman, L., Watson, M., Al Aqeel, A.I., Al-Aama, J.Y., Al-Mulla, F., Alonso, S., Aretz, S., Auerbach, A.D., Bapat, B., Bernstein, I.T., Bhak, J., Bleoo, S.L., Blöcker, H., Brenner, S.E., Burn, J., Bustamante, M., Calzone, R., Cambon-Thomsen, A., Cargill, M., Carrera, P., Cavedon, L., Cho, Y.S., Chung, Y.J., Claustres, M., Cutting, G., Dalgleish, R., den Dunnen, J.T., Díaz, C., Dobrowolski, S., Dos Santos, M.R., Ekong, R., Flanagan, S.B., Flicek, P., Furukawa, Y., Genuardi, M., Ghang, H., Golubenko, M.V., Greenblatt, M.S., Hamosh, A., Hancock, J.M., Hardison, R., Harrison, T.M., Hoffmann, R., Horaitis, R., Howard, H.J., Barash, C.I., Izagirre, N., Jung, J., Kojima, T., Laradi, S., Lee, Y.S., Lee, J.Y., Gil-da-Silva-Lopes, V.L., Macrae, F.A., Maglott, D., Marafie, M.J., Marsh, S.G., Matsubara, Y., Messiaen, L.M., Möslein, G., Netea, M.G., Norton, M.L., Oefner, P.J., Oetting, W.S., O'Leary, J.C., de Ramirez, A.M., Paalman, M.H., Parboosingh, J., Patrinos, G.P., Perozzi, G., Phillips, I.R., Povey, S., Prasad, S., Qi, M., Quin, D.J., Ramesar, R.S., Richards, C.S., Savige, J., Scheible, D.G., Scott, R.J., Seminara, D., Shephard, E.A., Sijmons, R.H., Smith, T.D., Sobrido, M.J., Tanaka, T., Tavtigian, S.V., Taylor, G.R., Teague, J., Töpel, T., Ullman-Cullere, M., Utsunomiya, J., van Kranen, H.J., Vihinen, M., Webb, E., Weber, T.K., Yeager, M., Yeom, Y.I., Yim, S.H., Yoo, H.S., 2009. Planning the human variome project: the Spain report. Hum. Mutat. 30, 496−510.

Karageorgos, I., Giannopoulou, E., Mizzi, C., Pavlidis, C., Peters, B., Karamitri, A., Zagoriti, Z., Stenson, P., Kalofonos, H.P., Drmanac, R., Borg, J., Cooper, D.N., Katsila, T., Patrinos, G.P., 2015. Identification of cancer predisposition variants using a next generation sequencing-based family genomics approach. Hum. Genom. 9, 12.

Kleanthous, M., Patsalis, P.C., Drousiotou, A., Motazacker, M., Christodoulou, K., Cariolou, M., Baysal, E., Khrizi, K., Pourfarzad, F., Moghimi, B., van Baal, S., Deltas, C.C., Najmabadi, S., Patrinos, G.P., 2006. The Cypriot and Iranian national mutation databases. Hum. Mutat. 27, 598−599.

Landrum, M.J., Lee, J.M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., Jang, W., Katz, K., Ovetsky, M., Riley, G., Sethi, A., Tully, R., Villamarin-Salomon, R., Rubinstein, W., Maglott, D.R., 2016. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res. 44, D862−D868.

McKusick, V.A., 1966. Mendelian Inheritance in Man. A Catalog of Human Genes and Genetic Disorders, first ed. Johns Hopkins University Press, Baltimore, MD.

Mitropoulou, C., Webb, A.J., Mitropoulos, K., Brookes, A.J., Patrinos, G.P., 2010. Locus-specific databases domain and data content analysis: evolution and content maturation towards clinical use. Hum. Mutat. 31, 1109−1116.

Mons, B., Velterop, J., 2009. Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009). Nano-Publication in the e-science era, Washington, DC, USA.

Mons, B., van Haagen, H., Chichester, C., Hoen, P.B., den Dunnen, J.T., van Ommen, G., van Mulligen, E., Singh, B., Hooft, R., Roos, M., Hammond, J., Kiesel, B., Giardine, B., Velterop, J., Groth, P., Schultes, E., 2011. The value of data. Nat. Genet. 43, 281−283.

Papadopoulos, P., Viennas, E., Gkantouna, V., Pavlidis, C., Bartsakoulia, M., Ioannou, Z.M., Ratbi, I., Sefiani, A., Tsaknakis, J., Poulas, K., Tzimas, G., Patrinos, G.P., 2014. Developments in FINDbase worldwide database for clinically relevant genomic variation allele frequencies. Nucleic Acids Res. 42, D1020−D1026.

Patrinos, G.P., 2006. National and ethnic mutation databases: recording populations' genography. Hum. Mutat. 27, 879−887.

Patrinos, G.P., Wajcman, H., 2004. Recording human globin gene variation. Hemoglobin 28, v−vii.

Patrinos, G.P., Brookes, A.J., 2005. DNA, disease and databases: disastrously deficient. Trends Genet. 21, 333−338.

Patrinos, G.P., Petricoin, E.F., 2009. A new scientific journal linked to a genetic database: towards a novel publication modality. Hum. Genom. Proteom. 1, 597478.

Patrinos, G.P., Giardine, B., Riemer, C., Miller, W., Chui, D.H., Anagnou, N.P., Wajcman, H., Hardison, R.C., 2004. Improvements in the HbVar database of human hemoglobin variants and thalassemia mutations for population and sequence variation studies. Nucleic Acids Res. 32, D537−D541.

Patrinos, G.P., Al Aama, J., Al Aqeel, A., Al-Mulla, F., Borg, J., Devereux, A., Felice, A.E., Macrae, F., Marafie, M.J., Petersen, M.B., Qi, M., Ramesar, R.S., Zlotogora, J., Cotton, R.G., 2011. Recommendations for genetic variation data capture in emerging and developing countries to ensure a comprehensive worldwide data collection. Hum. Mutat. 32, 2−9.

Patrinos, G.P., van Baal, S., Petersen, M.B., Papadakis, M.N., 2005a. The hellenic national mutation database: a prototype database for mutations leading to inherited disorders in the hellenic population. Hum. Mutat. 25 (4), 327−333.

Patrinos, G.P., Kollia, P., Papadakis, M.N., 2005b. Molecular diagnosis of inherited disorders: lessons from hemoglobinopathies. Hum. Mutat. 26, 399–412.

Patrinos, G.P., Cooper, D.N., van Mulligen, E., Gkantouna, V., Tzimas, G., Tatum, Z., Schultes, E., Roos, M., Mons, B., 2012b. Microattribution and nanopublication as means to incentivize the placement of human genome variation data into the public domain. Hum. Mutat. 33, 1503–1512.

Potamias, G., Lakiotaki, K., Katsila, T., Lee, M.T., Topouzis, S., Cooper, D.N., Patrinos, G.P., 2014. Deciphering next-generation pharmacogenomics: an information technology perspective. Open Biol. 4 pii:140071.

Riikonen, P., Vihinen, M., 1999. MUTbase: maintenance and analysis of distributed mutation databases. Bioinformatics 15, 852–859.

Ring, H.Z., Kwok, P.Y., Cotton, R.G., 2006. Human Variome Project: an international collaboration to catalogue human genetic variation. Pharmacogenomics 7, 969–972.

Scriver, C.R., 2004. After the genome - the phenome? J. Inherit. Metab. Dis. 27, 305–317.

Scriver, C.R., Nowacki, P.M., Lehvaslaiho, H., 1999. Guidelines and recommendations for content, structure, and deployment of mutation databases. Hum. Mutat. 13, 344–350.

Scriver, C.R., Waters, P.J., Sarkissian, C., Ryan, S., Prevost, L., Cote, D., Novak, J., Teebi, S., Nowacki, P.M., 2000. PAHdb: a locus-specific knowledgebase. Hum. Mutat. 15, 99–104.

Sipila, K., Aula, P., 2002. Database for the mutations of the Finnish disease heritage. Hum. Mutat. 19, 16–22.

Smith, T.D., Cotton, R.G., 2008. VariVis: a visualisation toolkit for variation databases. BMC Bioinforma. 9, 206.

Sosnay, P.R., Siklosi, K.R., Van Goor, F., Kaniecki, K., Yu, H., Sharma, N., Ramalho, A.S., Amaral, M.D., Dorfman, R., Zielenski, J., Masica, D.L., Karchin, R., Millen, L., Thomas, P.J., Patrinos, G.P., Corey, M., Lewis, M.H., Rommens, J.M., Castellani, C., Penland, C.M., Cutting, G.R., 2013. Defining the disease-liability of mutations in the cystic fibrosis transmembrane conductance regulator gene. Nat. Genet. 45, 1160–1167.

Soussi, T., 2000. The p53 tumor suppressor gene: from molecular biology to clinical investigation. Ann. N. Y. Acad. Sci. 910, 121–137.

Soussi, T., Beroud, C., 2001. Assessing TP53 status in human tumours to evaluate clinical outcome. Nat. Rev. Cancer 1, 233–240.

Soussi, T., Béroud, C., 2003. Significance of TP53 mutations in human cancer: a critical analysis of mutations at CpG dinucleotides. Hum. Mutat 21 (3), 192–200.

Soussi, T., Ishioka, C., Beroud, C., 2006. Locus-specific mutation databases: pitfalls and good practice based on the p53 experience. Nat. Rev. Cancer 6, 83–90.

Stein, L., 2002. Creating a bioinformatics nation. Nature 417, 119–120.

Stenson, P.D., Mort, M., Ball, E.V., Shaw, K., Phillips, A., Cooper, D.N., 2014. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Hum. Genet. 133, 1–9.

Thompson, B.A., Spurdle, A.B., Plazzer, J.P., Greenblatt, M.S., Akagi, K., Al-Mulla, F., Bapat, B., Bernstein, I., Capellá, G., den Dunnen, J.T., du Sart, D., Fabre, A., Farrell, M.P., Farrington, S.M., Frayling, I.M., Frebourg, T., Goldgar, D.E., Heinen, C.D., Holinski-Feder, E., Kohonen-Corish, M., Robinson, K.L., Leung, S.Y., Martins, A., Moller, P., Morak, M., Nystrom, M., Peltomaki, P., Pineda, M., Qi, M., Ramesar, R., Rasmussen, L.J., Royer-Pokora, B., Scott, R.J., Sijmons, R., Tavtigian, S.V., Tops, C.M., Weber, T., Wijnen, J., Woods, M.O., Macrae, F., Genuardi, M., InSiGHT, 2014a. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. Nat. Genet. 46, 107–115.

Thompson, R., Johnston, L., Taruscio, D., Monaco, L., Béroud, C., Gut, I.G., Hansson, M.G., 't Hoen, P.B., Patrinos, G.P., Dawkins, H., Ensini, M., Zatloukal, K., Koubi, D., Heslop, E., Paschall, J.E., Posada, M., Robinson, P.N., Bushby, K., Lochmüller, H., 2014b. RD-Connect: an integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research. J. Gen. Intern. Med. 29, 780–787.

Vail, P.J., Morris, B., van Kan, A., Burdett, B.C., Moyes, K., Theisen, A., Kerr, I.D., Wenstrup, R.J., Eggington, J.M., 2015. Comparison of locus-specific databases for *BRCA1* and *BRCA2* variants reveals disparity in variant classification within and among databases. J. Community Genet. 6, 351–359.

van Baal, S., Kaimakis, P., Phommarinh, M., Koumbi, D., Cuppens, H., Riccardino, F., Macek Jr., M., Scriver, C.R., Patrinos, G.P., 2007. FINDbase: a relational database recording frequencies of genetic defects leading to inherited disorders worldwide. Nucleic Acids Res. 35, D690–D695.

van Baal, S., Zlotogora, J., Lagoumintzis, G., Gkantouna, V., Tzimas, I., Poulas, K., Tsakalidis, A., Romeo, G., Patrinos, G.P., 2010. *ETHNOS*: a versatile electronic tool for the development and curation of national genetic databases. Hum. Genom. 4, 361–368.

Viennas, E., Gkantouna, V., Ioannou, M., Georgitsi, M., Rigou, M., Poulas, K., Patrinos, G.P., Tzimas, G., 2012. Population-ethnic group specific genome variation allele frequency data: a querying and visualization journey. Genomics 100, 93–101.

Wildeman, M., van Ophuizen, E., den Dunnen, J.T., Taschner, P.E., 2008. Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker. Hum. Mutat. 29, 6–13.

Zaimidou, S., van Baal, S., Smith, T.D., Mitropoulos, K., Ljujic, M., Radojkovic, D., Cotton, R.G., Patrinos, G.P., 2008. A_1ATVar: a relational database of human *SERPINA1* gene variants leading to alpha1-antitrypsin deficiency. Hum. Mutat.

Zlotogora, J., van Baal, S., Patrinos, G.P., 2007. Documentation of inherited disorders and mutation frequencies in the different religious communities in Israel in the Israeli National Genetic Database. Hum. Mutat. 28, 944–949.

Zlotogora, J., van Baal, J., Patrinos, G.P., 2009. The Israeli national genetics database. Isr. Med. Assoc. J. 11, 373–375.

Chapter 21

Molecular Diagnostic Applications in Forensic Science

H.E. McKiernan[1] and P.B. Danielson[1,2]
[1]Center for Forensic Science Research and Education, Willow Grove, PA, United States; [2]University of Denver, Denver, CO, United States

21.1 INTRODUCTION

Forensic DNA testing, first introduced in the mid-1980s (Gill et al., 1985; Jeffreys et al., 1985, 1992; Jeffreys, 1985), has aided the criminal justice system by assisting with the conviction of perpetrators and exonerating the innocent. Remains from missing persons and mass disasters have been identified through genetic testing, and parentage and kinship questions have been answered (Clayton et al., 1995; Prinz et al., 2007). New technologies and molecular applications are continually being embraced by the forensic biology community, resulting in regular and substantial improvements in the field's ability to generate information and to meet the throughput demands of law enforcement investigators and the judicial system. Advances that now allow for the cost-effective sequencing of entire human genomes and the development of instrument platforms that can sequence single molecules of DNA have already had a transformative impact on the biomedical community. The forensic community now stands to be transformed by these advances as well. This chapter will provide a broad overview of some of the history as well as the current and future directions of molecular-level forensic biological analyses.

21.1.1 Historic Overview of Human Identity Testing Prior to DNA

Historically, polymorphic protein markers held promise for human identity testing (Li, 2015). The first genetic tool used to compare individuals was the ABO blood types, which included four possible phenotypes: A, B, AB, and O. This blood group system was first discovered by Karl Landsteiner in the early 1900s and relied on the detection of antigenic variants on the surface of red blood cells (erythrocytes) (Yamamoto et al., 1990). While an inconsistency, or nonmatch, between an item of evidence and a reference sample was a powerful elimination tool, a match did not carry a high power of discrimination. This was unavoidable given that 44% of the US population has type O blood, and 40% has type A blood. Other genetic blood typing systems, such as the MN (named, respectively, after the second and fifth letters of the word immune) and Rh (named for rhesus monkeys where this antigen was first found) systems, were used for forensic applications (Bargagna and Pereira, 1967; Hagins et al., 1978; Springer and Desai, 1975). However, due to the limitations of these basic blood group systems, additional inherited protein polymorphic markers were assessed in an effort to increase the power of discrimination for human identity purposes. Variations in amino acid sequences responsible for protein variants could be readily identified by differences in molecular mass using standard electrophoretic separation and/or differences in molecular charge using isoelectric focusing electrophoresis (Budowle and Murch, 1986). Variants of phosphoglucomutase, a metabolic enzyme that facilitates the interconversion of glucose 1-phosphate and glucose 6-phosphate, were first described in the 1960s. The identification of specific isoforms was applied to the testing of both blood and semen samples. Human erythrocytes and blood serum contained a number of additional protein markers that featured different isoforms. The most common of these included erythrocyte acid phosphatase, esterase D, adenylate kinase, glyoxalase I, and adenosine deaminase (Li, 2015) (Table 21.1). While the combination of these protein markers improved discrimination between humans, each of these proteins typically contained only two or three isoforms. Thus even when used in conjunction with the blood typing systems, the power of discrimination failed to

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00021-3
Copyright © 2017 Elsevier Ltd. All rights reserved.

TABLE 21.1 Protein Markers Used for Pre-DNA Analysis of Blood and Semen Stains

Erythrocyte Isoenzyme	Protein Function	Possible Phenotypes
Adenosine deaminase (ADA)	Activates lymphocytes for immune response	ADA-1; ADA-2; ADA-2-1
Adenylate kinase (AK)	Provides energy state homeostasis	AK-1; AK-2; AK-2-1
Erythrocyte acid phosphatase (ACP/EAP)	Cleavage of phosphoryl groups during digestion	P^a; P^b; P^c
Esterase D (ESD)	Serine hydrolase	ESD-1; ESD-2; ESD-2-1
Glyoxalase I (GLO)	Detoxifies aldehyde-containing metabolic products	GLO-1; GLO-2; GLO-2-1
Phosphoglucomutase (PGM)	Metabolic enzyme in glycogenolysis and glycogenesis	PGM-1; PGM-2; PGM-2-1

even approach what is achievable with DNA-based profiling. Aside from their relatively poor power of discrimination, the use of protein-based identification systems also suffered from persistence issues due to adverse environmental conditions that are often encountered in a forensic context. Detection sensitivity presented still another challenge, as some protein markers were not present at reliably detectable levels in all tissue types. The ability to reproducibly type DNA from biological evidence, therefore, has been hailed as one of the most important developments in the field of forensic science.

21.1.2 DNA Testing: Development and Advantages

DNA profiling, as it is known today, evolved from two key breakthroughs in molecular biology that occurred around the same time. The first was in 1984 when University of Leicester geneticist Dr. Alec Jeffreys used variable number of tandem repeats and restriction fragment length polymorphism (RFLP) analysis of DNA to demonstrate genetic linkages between related individuals. In 1985, Dr. Jeffreys developed a "genetic profiling" method specifically for forensic use. The first application of this novel tool was as part of the investigation of the serial rape and murder of two young girls in the Enderby and Narborough villages near Leicestershire, Great Britain. A suspect, Richard Buckland, was arrested and confessed to the second murder. Genetic profiling, however, revealed that the DNA fingerprint of the semen from the two murdered girls did not match Mr. Buckland, who was found not guilty. A subsequent "DNA dragnet" was used in an effort to determine which, if any, of 4582 local men might have been the source of the semen from the two murdered girls. Although no match was found, a local baker named Colin Pitchfork was overheard boasting that he had gotten his friend to provide a DNA sample in his place—a boast that was pivotal to solving the crime (Gill et al., 1985; Jeffreys, 1985; Jeffreys et al., 1992; Wyman and White, 1980).

The second key breakthrough in molecular biology occurred across the Atlantic Ocean, where in 1985 chemist Dr. Kary Mullis developed the polymerase chain reaction (PCR) while working at the Cetus Corporation, a biotechnology firm in Emeryville, California. This elegantly simple technique would revolutionize forensic DNA profiling by making it possible to amplify specific regions of highly polymorphic DNA. This not only provided forensic analysts with the sensitivity necessary to obtain useful DNA profiles from challenging forensic samples, but it has also served as the foundation for the multiple generations of DNA-based typing methods used in forensic laboratories today.

In this early era of DNA analysis, several methods for RFLP analysis were employed, but all suffered from limited sensitivity. Even after that hurdle was overcome with the introduction of PCR-based assays, the availability of only a handful of genetic markers for PCR analysis meant that the power of discrimination was still inferior to that of the earlier RFLP methods. The Human Genome Project, which was formally launched in 1990, led to the modern era of forensic DNA profiling. The sequence generated from the Human Genome Project facilitated the discovery of a broad variety of short tandem repeat (STR) sequences. These highly polymorphic genetic markers were easily amplified by PCR-based assays, which facilitated population studies to assess their genetic variation. By the mid-1990s, forensic DNA testing had effectively combined a high degree of sensitivity afforded by PCR-based methods; a higher power of discrimination than had previously been possible; and greater throughput by the simultaneous amplification of multiple STR markers in a single multiplexed reaction. Since then, the field has experienced continual growth in the number of STR markers and classes of markers used. These include autosomal STRs, Y chromosome STRs, single nucleotide polymorphisms (SNPs), indels, and mitochondrial DNA (mtDNA) hypervariable region sequencing. Fig. 21.1 provides a historical overview of identity testing methodologies and developments prior to and after the advent of DNA

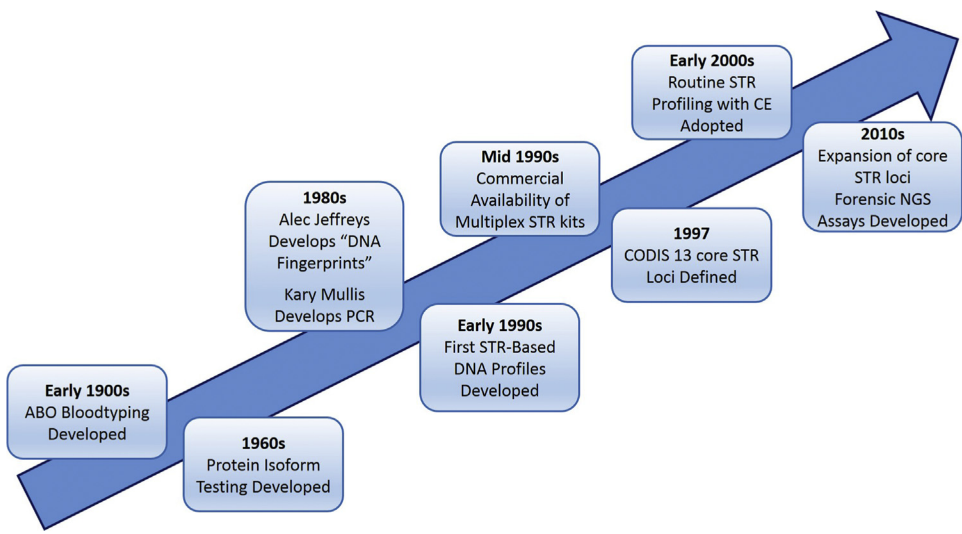


FIGURE 21.1 Historical overview of major 20th- and 21st-century advances in forensic identity testing. Prior to the advent of DNA profiling, forensic testing of blood and semen stains focused on protein biomarkers. Since the initial development of DNA profiling, forensic testing has emphasized the expansion of genetic marker systems and increased sample throughput. *CE*, capillary electrophoresis; *CODIS*, Combined DNA Index System; *NGS*, next-generation sequencing; *PCR*, polymerase chain reaction; *STR*, short tandem repeat.

profiling. Beyond this, forensic applications of next-generation sequencing (NGS) technologies are already emerging above the technology horizon. As has already been seen in the clinical diagnostics realm, the adoption of NGS assays has the potential to exponentially increase the amount of information that can be obtained from evidentiary samples. The wealth of sequence content provided by NGS will not only facilitate an even greater power of discrimination and the accurate resolution of complex DNA mixtures but may also assist in routine forensic analyses and provide investigative leads by providing useful information on an individual's kinship, ancestry, and even phenotype.

21.2 GENETIC MARKERS COMMONLY USED FOR FORENSIC ANALYSIS

The quantity and quality of DNA in biological samples are critical factors in determining which DNA profiling approach to take during forensic investigations. Evidentiary material may contain DNA from just one individual or a mixture of DNA from multiple individuals. Mixtures may be from individuals of the same or different genders. The potential presence and coextraction of PCR inhibitors present in soil, blood, tobacco, the dyes in clothing, and a host of other chemical agents are frequent concerns. Similarly, the exposure of biological materials to weathering, microbial action, ultraviolet radiation, or other environmental insults may result in DNA that is partially to severely degraded and thus refractory to amplification. Concern over the potential impact of these factors often dictates the choice of genetic markers that are employed by the forensic analyst in an effort to obtain potentially probative information from biological evidence. Depending on the type and condition of the forensic sample to be used, analysts can choose from a wide range of commercially available DNA typing systems using different genetic markers.

21.2.1 Short Tandem Repeat Markers

Microsatellites are a class of genetic markers that have proven to be particularly useful for forensic applications. Encompassing simple sequence repeats, expanded simple tandem repeats, and STRs, these markers consist of 2–9 base pair DNA motifs that are tandemly repeated, generally between 5 and 50 times (Tumpenny and Ellard, 2005). Microsatellites occur at thousands of locations throughout both human and nonhuman genomes and are notable for their relatively high mutation rates due to polymerase slippage. As a result, these markers are highly polymorphic such that the number of tandem repeat units varies widely among individuals in nearly all populations. Due to their small size and highly polymorphic nature, STRs have become the primary genetic marker used to individualize evidentiary material for forensic purposes.

Thousands of microsatellite loci have been characterized in the human genome (Collins et al., 2003; Edwards et al., 1991; Ellegren, 2004; Subramanian et al., 2003). This high abundance of STR loci provides a wide choice for incorporation into commercial kits manufactured for use by forensic laboratories. Analyzing amplified STRs entails separating various size fragments, usually by capillary electrophoresis, and detecting fluorescently labeled products (Fig. 21.2). Profiles generated from items of evidence

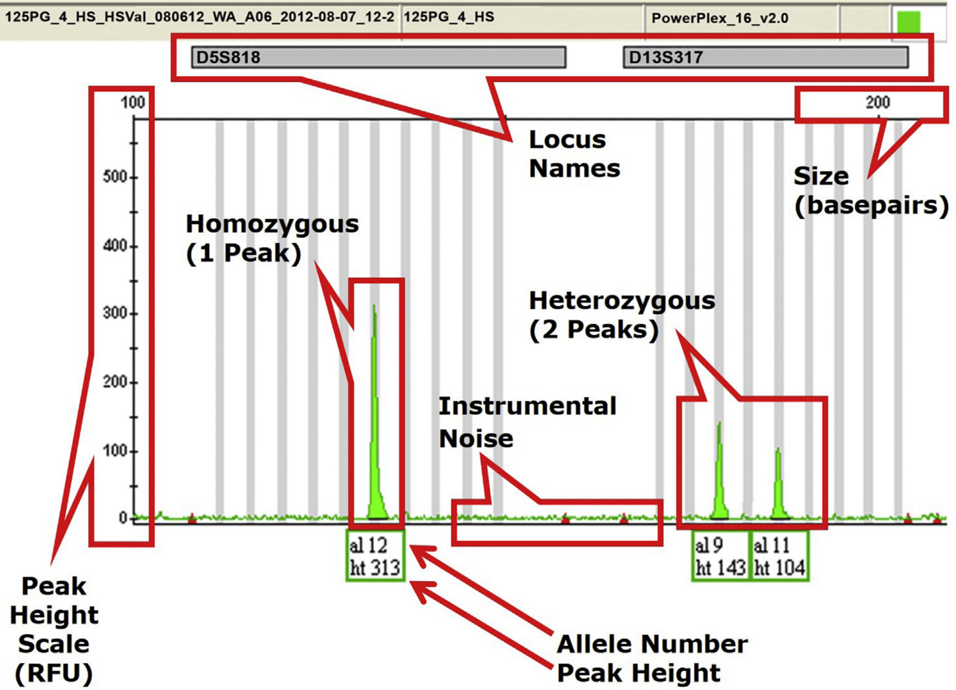


FIGURE 21.2 Illustrative example of a portion of a DNA profile. Shown are the DNA peaks in the analyzed data that were detected for 2 of the 13 core STR loci, D5S818 (homozygous) and D13S317 (heterozygous). Raw electrophoretic data were analyzed using GeneMapper® ID software, which labels each peak that is consistent with human DNA with an allele number and a peak height measured in relative fluorescence units. In general, a signal is readily discernible from instrument noise when it reaches twice the intensity of the highest instrument noise. *RFU*, relative fluorescence unit.

can then be compared to reference samples, and an interpretation of match, exclusion, or inconclusive can be made. The most useful STR loci, and those used routinely in the field, are those that exhibit sufficient variability among individuals to facilitate an inferred association of an evidentiary DNA profile to a specific individual of interest. These STR loci are considered to have a high power of discrimination, particularly when multiple STR loci are combined. Typically, the optimal loci for human identification purposes are those that exhibit the greatest variation in tandem repeats within human population groups.

One significant limitation of using STRs as genetic markers is that they suffer from varying amounts of a phenomenon known as stutter (Walsh et al., 1996). Stutter is an amplification artifact that is created during the PCR process as a result of strand slippage. This causes the DNA polymerase to either fail to copy one or more repeating units or to copy the same repeating unit more than once. The former results in an amplification product that is one or more repeat units smaller in size than the true genomic allele while the latter results in an amplification product that is typically one repeat greater in size than the true genomic allele (Walsh et al., 1996). With shorter core sequence motif lengths (i.e., di- and trinucleotide repeats), the amount of stutter amplicon produced relative to the amount of the true allele produced (i.e., the stutter percentage) is so great that it can complicate an accurate discrimination between genomic alleles and stutter artifacts and thus the subsequent interpretation of DNA profiling results. STRs, with a core sequence motif length of four base pairs (i.e., tetranucleotide repeats), are associated with relatively low stutter percentages and thus have become the most widely used markers for human identification purposes (Edwards et al., 1991; Zhong et al., 1999). Additionally, tetranucleotide repeat loci allow for better resolution of alleles during electrophoretic separation, which generally allows for single base resolution among fragments that are under 500 base pairs in total length. Slightly larger penta- and hexanucleotide repeats have also been evaluated for their utility as genetic markers. While these markers have the advantage of being associated with little, if any, stutter artifacts, they are less common in the human genome, and their amplification is less robust than the shorter STRs (Bacher et al., 1999; Bacher and Schumm, 1998). As a result, it is the tetranucleotide repeats that have been shown to be best suited for forensic type samples where DNA degradation due to environmental and/or chemical insult often impedes the ability to efficiently amplify an interpretable genetic profile from larger sized genetic markers.

Early on, the need for a standardized set of core STR loci to facilitate the comparison of DNA profiles among different laboratories was recognized by the forensic community. In 1997, the US Federal Bureau of Investigation (FBI) Laboratory evaluated and selected 13 STR loci

(CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, and D21S11) as the core STR loci for forensic purposes. The combination of these 13 loci provides for a high power of discrimination with the random match probability (RMP) of an average profile of less than one in a trillion unrelated individuals. These 13 loci serve as the foundational points of comparison for the databases of DNA profiles contributed by federal, state, and local participating forensic laboratories that are linked to the Combined DNA Index System (CODIS). CODIS is the FBI's program of support for criminal justice DNA databases as well as the software that has been developed to manage these databases (Budowle et al., 1998). Similar sets of core STR loci have been selected by governmental entities for use outside of the United States. These include the European Standard Set (FGA, TH01, VWA, D1S1656, D2S441, D3S1358, D8S1179, D10S1248, D12S391, D18S51, D21S11, and D22S1045); the UK core loci (FGA, TH01, VWA, D2S1338, D3S1358, D8S1179, D16S539, D18S51, D19S433, D21S11); the German core loci (FGA, TH01, SE33, VWA, D3S1358, D8S1179, D18S51, D21S11); and the Interpol Standard Set (FGA, TH01, VWA, D3S1358, D8S1179, D18S51, D21S11. In addition, the non-STR locus Amelogenin is typically coamplified with the core STR loci as a sex-typing marker.

Multiple manufacturers now produce STR kits for US and international markets that include these core STR loci (Zhong et al., 1999). Originally, however, no single commercially available multiplex was available to amplify all 13 of the CODIS core STR loci. Thus two multiplex amplification reactions were required to generate a full DNA profile. The Promega Corporation achieved this with the combination of the PowerPlex® 1.1 and 2.1 kits while Applied Biosystems used a combination of the AmpFlSTR® Profiler Plus and COfiler kits. Since 2000, both of these companies have manufactured expanded multiplex kits, such as the AmpFlSTR® Identifiler® and PowerPlex®16, that allow for the simultaneous amplification of all 13 CODIS core STR loci along with two additional noncore STR loci and the sex-typing marker Amelogenin in a single multiplex reaction (Fig. 21.3). The improved amplification chemistries (e.g., optimized buffer formulations, modifications to PCR primers, and alternate polymerases) associated with these newer kits increased PCR performance, which facilitated the successful amplification of complete profiles from more challenging types of forensic samples, particularly those containing inhibitory compounds such as heme, humic acid, and tannic acid (Ensenberger et al., 2010). Based on extensive validation studies, the AmpFlSTR® Identifiler® and PowerPlex®16 kits were also shown to be at least twice as sensitive as the earlier generation AmpFlSTR® Profiler Plus and COfiler kits. Accordingly, these newer kits reproducibly yielded complete and interpretable DNA profiles with more challenging types of forensic samples, even those containing as little as 125 pg of human genomic DNA (equivalent to approximately 20 human somatic cells). Partial profiles were often generated from trace samples containing as little as 31 pg of DNA (equivalent to approximately five human cells).

In addition to the greater sensitivity offered by such "newer generation" kits, the analysis of low-template and challenging DNA samples has also benefited from the development of reduced-size STR kits such as the AmpFlSTR® MiniFiler kit. In cases where the quantity of DNA is severely limited (generally less than 100 pg) or the integrity of the DNA is substantially compromised (highly degraded), standard STR testing often produces DNA profiles characterized by allele and locus dropout at the larger STR loci, resulting in the detection of only a partial DNA profile. Partial DNA profiles are typically more difficult to interpret and generally do not provide the power of discrimination necessary to include or exclude potential contributor(s) with a high degree of confidence. The use of mini-STR amplicons makes it possible to overcome these challenges in many cases. Mini-STR kits employ primer pairs that preferentially target the larger STR loci with annealing sites that encompass less of the flanking sequence adjacent to the repetitive element, resulting in a smaller PCR product. The smaller size of the STRs favors efficient DNA amplification, resulting in improved detection sensitivity relative to that of standard STR kits. Overall, mini-STR kits allow for a more successful analysis of compromised and degraded samples.

In 2010, the FBI Laboratory began to investigate the utility of expanding the 13 core STR loci to include additional STR markers (Ge et al., 2012). Seven additional STR loci (D1S1656, D2S441, D2S1338, D10S1248, D12S391, D19S433, and D22S1045) were ultimately selected for the expansion. These additional genetic markers are to be integrated into the CODIS database by 2017. The purpose of this expansion in the number of core STR loci used for forensic DNA profiling is 2-fold. First, it will increase international concordance, thereby making it possible to better compare evidentiary and reference samples generated in different countries around the world. Second, it will increase the power of discrimination that can be attained when comparing profiles. This will reduce the possibility of an adventitious match, which is a concern that has been raised as national forensic databases, consisting of millions to tens of millions of DNA profiles, have grown. These seven additional STR loci have been incorporated into both the AmpFlSTR® GlobalFiler® kit from Thermo Fisher Scientific as well as the PowerPlex® Fusion kit from the Promega Corporation. These kits also enable the genotyping of additional genetic markers beyond those making up the expanded core STR loci. The AmpFlSTR® GlobalFiler® kit added SE33 and a Y indel locus while the

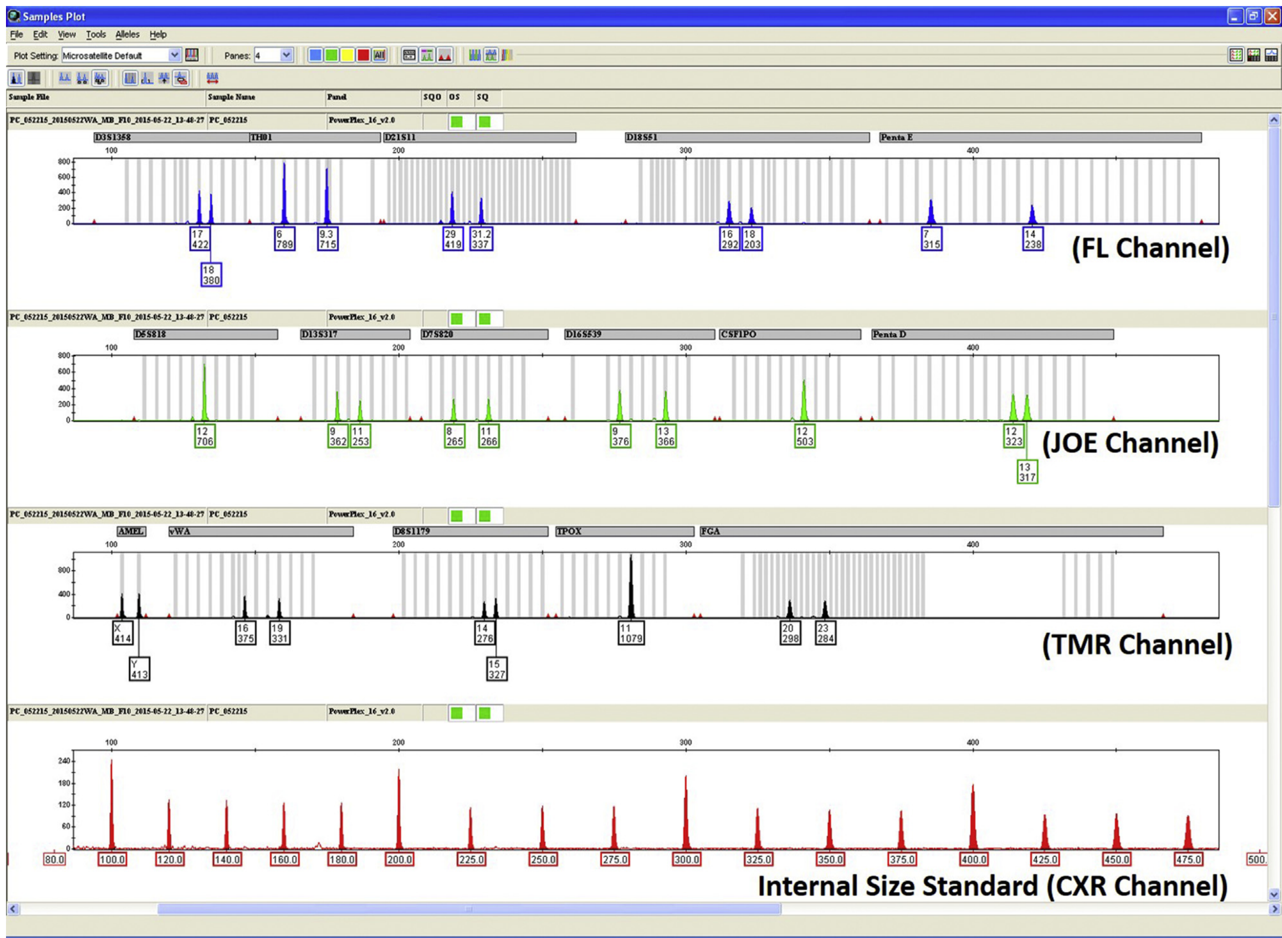


FIGURE 21.3 Illustration of a full DNA profile generated using the PowerPlex® 16 kit, which allows for multiplexed coamplification and three-color detection of 16 loci (13 core short tandem repeat loci, 2 pentanucelotide repeats, and Amelogenin). Amplified genetic markers in the blue channel are labeled with fluorescein; in the green channel with 6-carboxy-4′,5′-dichloro-2′,7′-dimethoxy-fluorescein; and in the black channel with carboxy-tetramethylrhodamine. The internal lane standard used to size the amplified genetic markers is labeled with carboxy-X-rhodamine, which is detected as a fourth color channel (*red* dark gray in print versions). *CXR*, carboxy-X-rhodamine; *FL*, fluorescein; *JOE*, 6-carboxy-4′,5′-dichloro-2′,7′-dimethoxy-fluorescein; *TMR*, carboxy-tetramethylrhodamine.

PowerPlex® Fusion kit added Penta D and Penta E. Both kits also include Amelogenin for gender determination and the Y chromosome STR locus DYS391 to facilitate linkage to Y chromosome haplotypes.

21.2.2 Y Chromosome Markers

Even with their enhanced sensitivity and ability to overcome inhibitors present in a sample, modern STR kits such as the AmpFlSTR® GlobalFiler® and PowerPlex® Fusion kits (like their forerunners) do not have the ability to selectively amplify and obtain an interpretable male DNA profile from intimate samples such as oral, vaginal, and rectal smears and/or swabs or any other evidentiary items where an excess of female DNA relative to male DNA would typically make it difficult to impossible to detect the male contributor. This is because when standard (i.e., autosomal) STR kits are used for human DNA profiling, the amplification and detection of a male DNA profile can be easily masked in a mixture by the presence of more than a 10- to 20-fold excess of female DNA. It is for these types of samples that kits restricted to Y chromosome markers are useful.

The Y chromosome is the smallest of the human chromosomes, encompassing only about 60 million base pairs of sequence. Y chromosome DNA, as with standard STR markers, resides in the nucleus; however, there are some notable differences between markers that reside on the Y chromosome with respect to autosomal STR loci (Gill et al., 2001). First and foremost, the Y chromosome is inherited paternally, and nearly all of the DNA in the Y chromosome is nonrecombinant, except for the most distal portions of the Y chromosome (Skaletsky et al., 2003). Therefore profiles generated from Y chromosome markers are passed from one generation to the next as a

stable haplotype. As a result, a Y haplotype is typically identical (apart from instances of relatively rare meiotic mutation) among all of the male relatives of a given paternal family lineage. This fact makes Y chromosome markers useful in human identity and paternity cases but greatly decreases the power of discrimination that is attainable relative to autosomal STR loci.

Y chromosome genetic markers include both biallelic SNPs (Y-SNPs) and an *Alu* element variant as well as over 200 characterized multiallelic STR (Y-STR) markers. In 1997, the International Y-STR User Group recommended the use of nine Y-STR loci (DYS19, DYS385a, DYS385b, DYS3891, DYS38911, DYS390, DYS391, DYS392, and DYS393) as the European minimal haplotype core set of Y-STR loci (Kayser, 2003; Schneider et al., 1998). In 2003, the Scientific Working Group on DNA Analysis Methods (SWGDAM) recommended the inclusion of two additional loci (DYS438 and DYS439) as core Y-STR loci for forensic applications in the United States (Ayub et al., 2000). Commercial amplification kits that could type all 11 markers included the PowerPlex® Y System kit from the Promega Corporation and the AmpFlSTR® Yfiler® kit from Applied Biosystems. Newer generations of commercially available Y-STR kits, including the PowerPlex® Y23 and AmpFlSTR® Yfiler Plus kits, have since been developed, which have multiplexed the amplification of 23 and 27 Y-STR loci, respectively. These kits combined the loci previously used with new loci that offer higher allelic diversities, allowing for improved distinction among unrelated male individuals (i.e., a higher power of discrimination) that shared some of the more common haplotypes seen in the general population. Rapidly mutating Y-STR loci were also incorporated to facilitate discrimination between paternally related individuals.

In spite of its limitations with respect to the achievable power of discrimination, there are a number of significant advantages that male-specific approaches to DNA profiling offer in the context of forensic testing. These include allowing a forensic analyst to more readily detect the presence of potentially probative male DNA in samples previously thought to contain only female profiles. This is particularly relevant in any challenging cases (especially sexual assaults) where some items of evidence are likely to consist of DNA mixtures containing more than a 10- to 20-fold excess of female DNA. Additionally, paternal transmission of Y haplotypes allows for the tracking of family lineage, which can be applied to paternity testing, missing persons investigations, human migration and evolution studies, and historical and genealogical research (Butler, 2005).

21.2.3 Mitochondrial DNA Markers

Typically, polymorphic nuclear DNA testing (i.e., either autosomal or Y-STR) is used for comparative purposes in forensic investigations. This approach, however, often fails with items like aged bones and teeth or shed/broken hair shafts with little or no root tissue. This is due to the limited amounts or extensive degradation of high molecular weight nuclear DNA that characterizes these types of samples. When faced with such challenging samples, forensic analysts can turn to mtDNA analysis in an effort to obtain potentially useful genetic information.

The mtDNA genome consists of 16.5 kilobases. The two strands of the mtDNA genome (termed heavy and light) differ considerably in their base compositions. The heavy strand is rich in guanine, and the light strand is rich in cytosine. The mitochondrial genome contains very little noncoding sequence (7%) in contrast to the nuclear genome where approximately 97% is not expressed. The 37 genes (22 mitochondrial tRNAs, 2 mitochondrial rRNAs, and 13 mitochondrial respiratory complex subunits) account for approximately 93% of the mtDNA sequence. These genes lack introns, are nearly contiguous with each other, and even overlap in some cases. With the exception of the one or two noncoding bases that often exist between neighboring genes, the only significant region of the mitochondrial genome free of coding sequence is the displacement loop region, also known as the D-loop. A duplicated stretch of the heavy strand exists as an independent fragment (7S DNA). This fragment binds to the D-loop, where it participates in the formation of a triple helical DNA structure known as the "control region," which contains the origin of replication for the heavy strand and the predominant transcriptional promoters (Greenberg et al., 1983).

Several characteristics of the mitochondrial genome make it an attractive target for forensic analysis. First, the subcellular sequestration of the mtDNA within the mitochondrion and its circular nature contribute to its stability. Second, there can be four to five copies of the mtDNA genome residing in the cytoplasm of each mitochondria (Satoh and Kuroiwa, 1991) and hundreds of mitochondria per cell, depending on the cell type. In many tissues this results in a high mtDNA copy number per cell (i.e., hundreds to thousands of copies of the mtDNA genome per cell) as opposed to just one copy of the diploid nuclear genome (Bogenhagen and Clayton, 1974). This feature facilitates the successful analysis of highly degraded and/or very limited amounts of starting material (Allen et al., 1998; Robin and Wong, 1988). Third, the mitochondrial genome is maternally inherited by both males and females, and thus an individual's mtDNA haplotype is a direct reflection of the up to 100,000 copies of the mitochondrial genome present in the oocyte at fertilization. This makes the mtDNA haplotype the female lineage equivalent of the Y-STR haplotype. The maternal transmission of mtDNA-haplotypes allows for the tracking of family lineage, which can be applied to parentage testing, missing persons

investigations, human migration and evolution studies, and historical and genealogical research.

In contrast to the majority of forensic DNA profiling methods that rely on the identification of length variants at STR loci in nuclear DNA, mtDNA profiling relies on the direct examination of DNA sequence variants relative to the Revised Cambridge Reference Sequence (Andrews et al., 1999). Portions of the D-loop region designated as hypervariable region 1 (HV-1) and hypervariable region 2 (HV-2) are of particular forensic utility because of the high degree of sequence variants they exhibit within human populations. Accordingly, the direct sequencing of mtDNA has been used to identify skeletal remains from the Vietnam War (Holland et al., 1993); to determine that bone material from a grave in Yekaterinburg, Russia likely originated from members of the imperial Romanov family (Gill et al., 1994; Ivanov et al., 1996); to identify shed head hairs and saliva from robbery caps (Allen et al., 1998); and to identify the skeletal remains of murder victims (Hagelberg et al., 1991).

Just as with Y-STRs, mtDNA does not undergo recombination. Thus genetic diversity is generated through mutation. By some estimates, mutations arise and are fixed in the mitochondrial genome at a rate approaching 10 times that of equivalent sequences in nuclear genomes. A number of factors have been proposed to account for this sequence instability, including an inefficient system for DNA repair, oxidative damage, and the greater number of replicative cycles that mtDNA undergoes relative to nuclear DNA during cell growth and development.

The preponderance of the evidence examined to date suggests that in most individuals, in spite of an accelerated mutation rate, the majority of mtDNA molecules within any individual at maturity will still be represented by a single sequence (i.e., homoplasmy). Occasionally, a de novo mutation may arise and proliferate within the mtDNA population, resulting in the existence of two distinct mtDNA haplotypes within a single individual (i.e., heteroplasmy). This can increase the power of discrimination of forensic mtDNA analysis by providing an additional character state for inclusion in the mtDNA haplotype. This was the case for the forensic analyses of mtDNA that was used to help identify the remains of Tsar Nicholas Romanov II, exhumed in 1991 from Yekaterinburg, Russia (Ivanov et al., 1996). Sequencing mtDNA from bone fragments suspected of being those of the Tsar revealed a heteroplasmic mixture of mtDNAs at position 16,169 in HV-1, where C and T accounted for about 70% and 30% of the mixture, respectively. Sequencing of mtDNA from the exhumed remains of the Tsar's brother (Grand Duke George Alexandrovich Romanov) revealed a similarly rare heteroplasmic mixture of C and T at position 16,169. When combined with sequence identity throughout the remainder of the D-loop mtDNA sequence, the inclusion of a shared heteroplasmy significantly strengthens the weight of the mtDNA match.

Aside from forensic applications, mtDNA genetics is extensively studied and applied to other fields, especially medical diagnostics. There have been significant advances made in understanding mtDNA mutations and their link to inherited diseases as well as in identifying acquired mtDNA mutations in aging and cancer (Taylor and Turnbull, 2005).

21.2.4 Single Nucleotide Polymorphism Markers

The newest genetic markers to be adopted by the forensic community are SNPs. These single nucleotide variants, usually biallelic in nature, can occur at any point in the genome. Originally arising via a spontaneous point mutation, some SNP loci have become fixed in certain populations due to evolutionary forces and genetic drift. The overall abundance of SNPs in the human genome, estimated to be millions, makes them a marker of forensic interest. Additionally, SNPs are typically detected as part of small amplicons that are typically only 45–55 base pairs in length (approximately the size of two PCR primers). This makes them ideal markers for degraded DNA samples. Aside from their utility in identity testing, SNPs can also be used to provide lineage information, identifying relatives based on their location in the mtDNA genome or the Y chromosome; ancestry information, indicating the global biogeographic origin of an individual; and phenotype information, such as skin, hair, and eye color for the generation of investigative leads.

Two research groups leading the field early on in SNP characterization for identity testing were Dr. Kenneth Kidd's research group at Yale University and the SNP*for*ID Consortium in Europe (Karafet et al., 2008; Pakstis et al., 2010; Phillips et al., 2007). Research from both groups was focused on identifying SNP markers with high heterozygosity to maximize the amount of information garnered from each SNP and a low fixation index to reduce chance differences between populations. Given their biallelic nature, however, individual SNPs have a relatively low power of discrimination compared to existing STR markers. Therefore in order to attain equivalent RMPs and to be useful for the individualization of a forensic DNA profile, a larger number of SNP loci must be multiplexed and accurately called. Depending on the specific panels of SNPs employed, approximately 25–42 SNP loci must be queried to achieve a statistical power of discrimination approaching that of the 13 core STR loci. In the past, this represented a significant technical obstacle to the adoption of SNPs by forensic labs. Advances in NGS technologies have done much to overcome this obstacle. Commercial products have since incorporated the most informative SNP loci identified by the Kidd research group and the SNP*for*ID Consortium.

These include the HID-Ion AmpliSeq Identity Panel by Thermo Fisher, which incorporates 124 autosomal markers, and the ForenSeq DNA Signature Prep Kit by Illumina, which, among other markers, targets 95 identity SNPs.

Genetically linked autosomal SNPs are inherited as a haplotype since linkage impedes recombination. Additionally, haplotype SNPs can be found on the Y chromosome and mtDNA genome as well as the X chromosome. As previously discussed with regard to Y chromosome and mtDNA haplotypes, and given the low mutation rates of SNP markers, SNPs can be useful for aiding multiple types of forensic cases involving kinship analysis. These low mutation rates are also what allow SNPs to become fixed in populations, which allows them to be used as indicators of ancestry and ethnicity.

Historically, DNA generated from an item of evidence was not useful without a known reference sample from a suspect for comparison. When that sample did not match any profiles in a DNA database, little more could be done with it. Current research, however, has identified phenotypic SNPs that can be used to generate investigative leads from the evidentiary material itself. SNPs correlating to genes for skin, hair, and eye color have been evaluated and can be used to develop beneficial information regarding the physical appearance of a victim or suspect in a case.

In diagnostic and clinical medical research, SNPs have also been identified that directly relate to human health status. For example, an individual's response to certain drugs can be determined by SNPs involved with cytochrome p450s. SNPs can also be used to determine an individual's risk of developing a particular disease or his or her susceptibility to toxins as well as tracking disease inheritance.

21.3 DNA EXTRACTION METHODOLOGIES

Direct PCR kits have allowed for DNA profiling directly from a buccal swab or bloodstain without the need for separate DNA extraction or quantitation steps. Such direct approaches are well suited to nonchallenging samples where biological material is abundant and pristine. This would include reference samples and database samples. For the majority of forensic casework-type samples, however, it is prudent and often necessary to separate the DNA from other cellular material and inhibitory compounds that may be present in the sample that could interfere with the downstream analysis of DNA. Given the multitude of sample types and matrices received for forensic testing, there are a number of options for DNA extraction. A sufficient quantity, quality, and purity of DNA for successful testing is the ultimate goal of any extraction technique. Additional considerations when selecting an extraction method include throughput and cost-effectiveness. In general, however, most extraction protocols involve the enzymatic digestion of tissue, cells, and subcellular organelles, the disassociation of DNA from nuclear proteins, and the removal of these nonnucleic acid components from the sample.

21.3.1 Organic (Phenol−Chloroform) Extraction

Organic (phenol−chloroform) extraction uses sodium dodecylsulfate (SDS) and proteinase K for the enzymatic digestion of proteins and nonnucleic acid cellular components (Fig. 21.4). A mixture of phenol:chloroform:isoamyl alcohol (25:24:1) is then added to promote the partitioning of lipids and cellular debris into the organic phase, leaving isolated DNA in the aqueous phase. Following centrifugation, the aqueous phase containing the purified DNA can be transferred to a clean tube for analysis. DNA can also be recovered and concentrated from the aqueous phase by ethanol precipitation or through the use of a centrifugal filter unit (i.e., Vivacon® or Amicon® devices), which allows for additional purification and concentration of the DNA in the samples (Koons et al., 1994). Organic extraction recovers double-stranded DNA and was required for early RFLP methods. While this method remains one of the most reliable and efficient, it is also very time-consuming, uses hazardous chemicals, and, because of the greater hands-on effort and multiple tube transfers involved, introduces increased opportunities for contamination and sample mishandling (Köchl et al., 2005).

21.3.2 Chelation Extraction

Extractions based on the use of chelating ion exchange resins were introduced to forensic laboratories in the early 1990s. Among the most popular of these resins was Chelex® from Bio Rad (Fig. 21.4). Consisting of a styrene divinylbenzene copolymer containing paired iminodiacetate ions, the resin is added directly to a sample where it acts to bind polyvalent metal ions such as magnesium (critical cofactors for the functionality of endogenous nuclease enzymes). Heat is applied to lyse the cells, releasing DNA, while the chelating ion exchange resin protects the DNA from degradation. Chelex® resin and cellular debris are then removed via centrifugation (Walsh et al., 1991). Chelation-based extractions are fast, can be easily automated, and require minimal sample transfer, which decreases the opportunity for contamination or other sample mishandling. A key disadvantage associated with the use of chelation extraction for the isolation of DNA is that the method provides no mechanism for the removal of inhibitory compounds (e.g., hematin and immunoglobulin gamma in whole blood or humic acid in soil-contaminated samples).

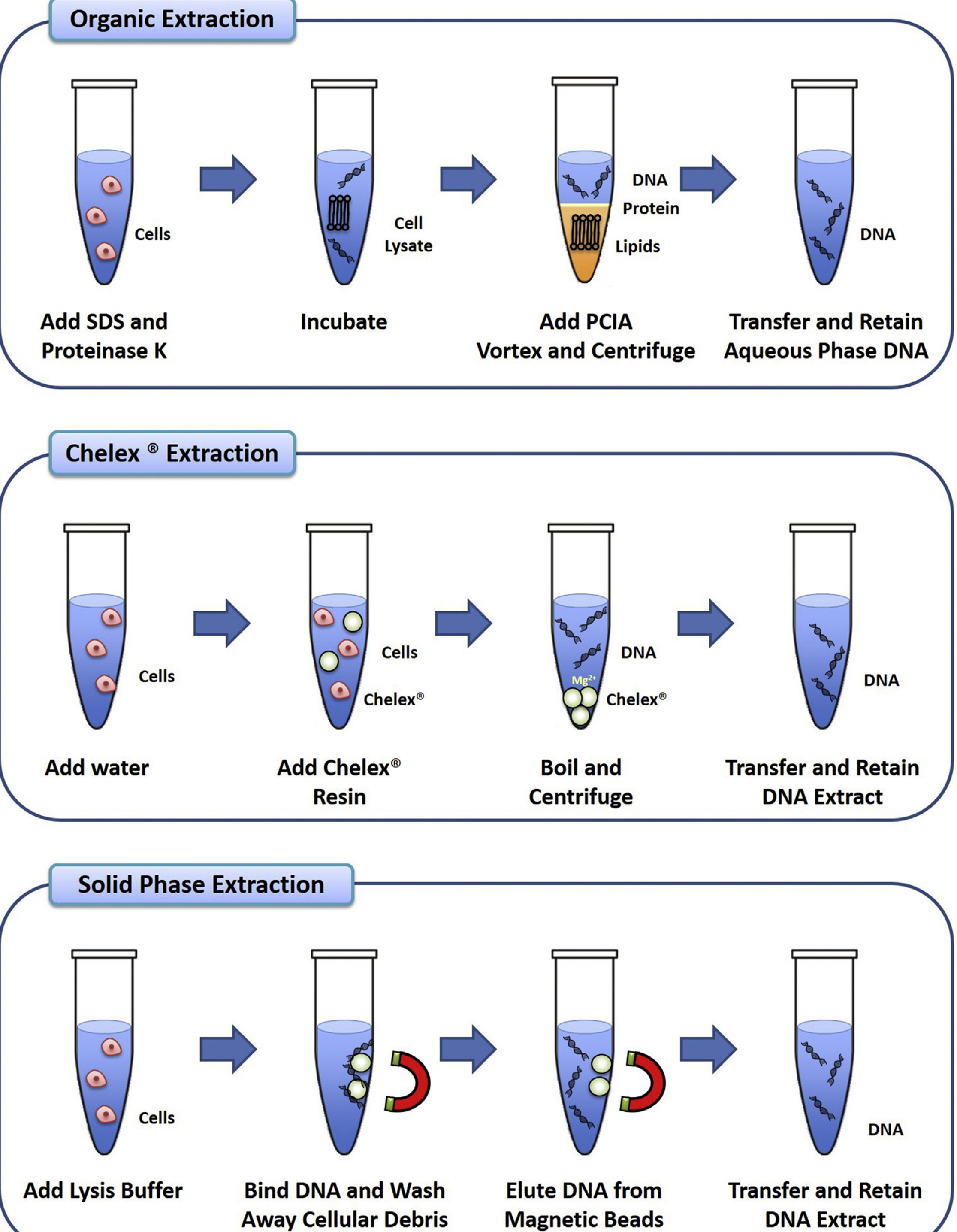


FIGURE 21.4 Illustration comparing three major methods of DNA extraction. Top: Organic extraction using phenol:chloroform:isoamyl alcohol partitions DNA to the aqueous phase while lipids and proteins are partitioned to the organic phase or the aqueous–organic interface. Organic extraction yields high-purity DNA. Middle: Extraction by chelating resin (e.g., Chelex®) releases DNA by boiling and inactivates degradative enzymes by chelating essential divalent cations such as Mg^{2+}. Chelex® extraction yields a lower purity DNA and does not remove polymerase chain reaction inhibitors. Bottom: Solid phase extraction selectively binds DNA to a solid support such as silica in the presence of high salt solutions. By coupling silica to paramagnetic particles, magnets can be used to immobilize DNA while washing away other cellular components. Solid phase extraction yields high-purity DNA and is readily implemented on automated platforms. *PCIA*, phenol:chloroform:isoamyl alcohol; *SDS*, sodium dodecylsulfate.

21.3.3 Solid Phase Extraction

More readily automatable extraction techniques involve solid phase extraction methods (Fig. 21.4). These techniques selectively bind DNA to a solid surface, such as silica in the presence of high concentrations of salts. The bound nucleic acids and associated substrates are then separated from the remaining cellular debris through the use of magnets or centrifugation. Purified DNA is then readily eluted from the solid surface by the immersion of low ionic strength or pH-adjusted buffers, allowing for nucleic acid recovery and concentration. Products such as QIAamp® spin columns from Qiagen selectively bind DNA to silica in the presence of chaotropic salts and elute DNA under alkaline conditions (Scarpetta et al., 1998). The DNA IQ kit from the Promega Corporation also uses a silica-based paramagnetic binding resin that allows for the use of magnets to isolate the silica-bound DNA on the side of a sample tube. This eliminates the need for tube transfers during the wash steps (Krnajski et al., 2007). The PrepFiler system by Thermo Fisher Scientific also uses a paramagnetic resin, but rather than silica, it employs a coating of a dextran derivative, which binds DNA in the presence of alcohol (Brevnov et al., 2009).

21.3.4 Differential Extraction

Differential extraction is a modified extraction technique allowing for the selective lysis and isolation of DNA from a mixture of sperm and epithelial cells (Fig. 21.5). Forensic samples, such as those from sexual assault kits, may contain male sperm cells mixed with male and female epithelial cells. Generally, it is the male sperm cells that are of greatest forensic value. Differential extraction techniques allow for the selective physical isolation of the male fraction prior to DNA profiling (Yoshida et al., 1995). During differential extraction, the epithelial cells are first selectively lysed through the addition of SDS and proteinase K. Sperm cells are generally resistant to lysis by these chemicals. Once the epithelial cells have been lysed, centrifugation is used to selectively pellet any intact sperm cells. The supernatant-containing DNA from male and female epithelial cells can then be removed to a separate tube. The isolated sperm cells are then lysed by the addition of SDS, proteinase K, and dithiothreitol (DTT). DTT reduces the disulfide bonds present in the sperm nuclear membranes, releasing sperm cell DNA. The DNA in the separate epithelial cell fraction and the sperm cell fraction can then be isolated by organic or nonorganic methods. This extraction method can simplify the interpretation of DNA

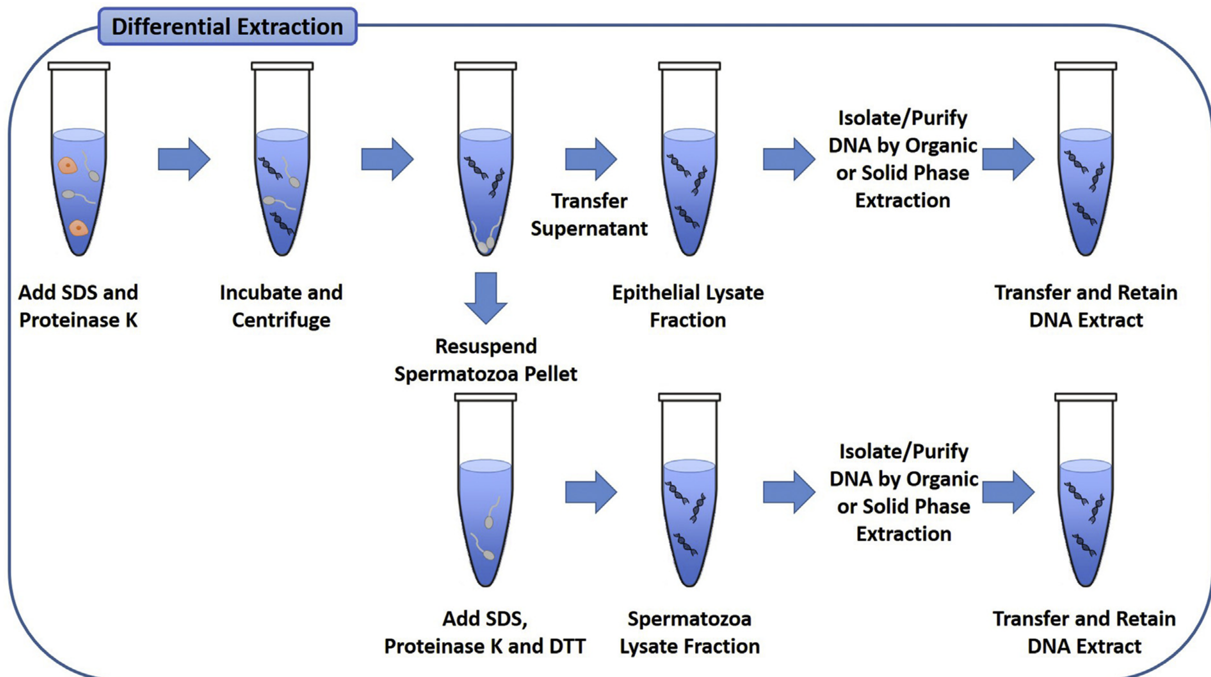


FIGURE 21.5 Illustration of the differential extraction workflow, which allows analysts to partition epithelial cell DNA from sperm cell DNA. This method of DNA extraction takes advantage of the protein disulfide bonds in the outer membrane of spermatozoa, which render them more resistant to lysis than epithelial cells. Commonly employed for the processing of sexual assault evidence, epithelial cells are first lysed, leaving intact spermatozoa, which are readily pelleted by centrifugation. The removal of the epithelial cell lysate to a separate tube leaves behind a pellet that is enriched for spermatozoa. The subsequent denaturation of protein disulfide bonds by dithiothreitol facilitates the lysis of the sperm cells for DNA extraction. *DTT*, dithiothreitol; *SDS*, sodium dodecylsulfate.

profiling results by increasing the likelihood of obtaining a clear male DNA profile from the sperm fraction. Since the epithelial fraction contains both male and female DNA, this may still present as a DNA mixture. It should be emphasized that this is a pre-PCR approach that physically isolates male cells for autosomal STR analysis, which offers a superior power of discrimination. This is in contrast to the use of Y-STRs, which simply use male targeted genetic markers to selectively amplify male DNA in the presence of excess female DNA.

21.4 DNA QUANTITATION

After the DNA in a biological sample has been isolated, the quantity and quality must be assessed so as to facilitate successful DNA profiling. This not only represents prudent laboratory practice, but also Standard 9.4 of the FBI's Quality Assurance Standards for Forensic DNA Testing Laboratories mandate a human-specific quantitation step for forensic samples (FBI, 2008). PCR-based assays, such as the multiplex STR assays described in Section 21.2, work optimally within a relatively narrow range of DNA input quantities. For most commercial STR kits, this is usually around 0.5−1.0 ng of input human DNA. If too much DNA is added for amplification, the resulting electropherograms often include off-scale allelic peaks (i.e., fluorescent signals from amplified alleles that exceed the dynamic range of the detector). Off-scale peaks are characterized by the presence of a split peak, with the depth of the split being correlated with the amount of detector saturation. While minor saturation may not have a substantial quantifiable impact on profile interpretation (e.g., stutter ratios), excessive saturation results in deep splits in the allelic peak, which can complicate the accurate quantitative interpretation of a DNA profile (Fig. 21.6). If too little DNA is added for amplification, there is a corresponding decrease in the average peak height of detected alleles. As DNA input quantities approach the sensitivity limits of the DNA profiling kits used, complete/full profiles give way to incomplete/partial profiles along with an increase in stochastic amplification phenomena such as allele dropout and heterozygote imbalance.

Commercial real-time quantification kits can accurately and precisely determine the amount of male and female DNA present in a sample as well as indicate whether or not a sample contains inhibitors or degraded DNA that could reduce the amplification efficiency or even completely block the ability to obtain an interpretable DNA profile. Thus data from DNA quantitation assays not only facilitates the use of optimal DNA input quantities, but it can also serve as a quality control step to guide analysts in choosing the most appropriate DNA profiling strategy, including the choice of genetic markers for DNA profiling and/or the need for additional preamplification purification/concentration.

21.4.1 Real-Time Quantitative Polymerase Chain Reaction

The most common approach for DNA quantitation employed by forensic laboratories is real-time PCR (qPCR) using a fluorogenic 5′ nuclease or TaqMan® assay. A TaqMan® probe is designed to be complementary to a specific DNA region of interest located between the forward and reverse PCR primers. The probe is labeled with a high-energy "reporter" dye on the 5′ end and a low-energy "quencher" dye on the 3′ end. When the probe is intact, the two dyes are in close proximity to each other, and efficient fluorescence resonance energy transfer occurs between them, quenching any fluorescence. During PCR elongation, the 5′ exonuclease activity of *Taq* DNA polymerase cleaves the TaqMan® probe, releasing the reporter dye, allowing fluorescence to be detected. Fluorescence intensity is measured and is proportional to the amount of target DNA synthesized during each PCR cycle. By quantifying the cycle-to-cycle change in fluorescence over multiple PCR cycles, it is possible to determine the concentration of DNA in the sample. The more DNA present in the original sample, the sooner a detectable level of fluorescence is achieved. This is measured in terms of PCR cycles. The cycle threshold (C_T) is the number of PCR cycles required to reach a predefined threshold of fluorescence. For absolute DNA quantitation, the log of the DNA quantities for a dilution series of a DNA standard of known concentration is plotted against the resulting C_T values for each dilution. This produces a linear standard curve, which can then be used to determine the amount of DNA in a forensic sample.

21.4.2 Commercial Quantitative Polymerase Chain Reaction Kits

Multiplexing the qPCR assay is possible by the use of different TaqMan® probes for multiple targets of interest, with each probe being labeled with a different reporter dye. In a basic qPCR kit for measuring total human DNA only, there are typically two TaqMan® probes. One is targeted to a human-specific gene for quantification purposes, and the other is targeted to a synthetic oligonucleotide sequence, which is present in the reaction mix at a known concentration. This synthetic oligonucleotide, for which the C_T value is known a priori, serves as an internal positive control (IPC). Deviation from the expected C_T value for the IPC (i.e., an increase in the C_T value) indicates the presence of PCR inhibitors in a sample. This can provide guidance to a forensic analyst by identifying samples for which additional

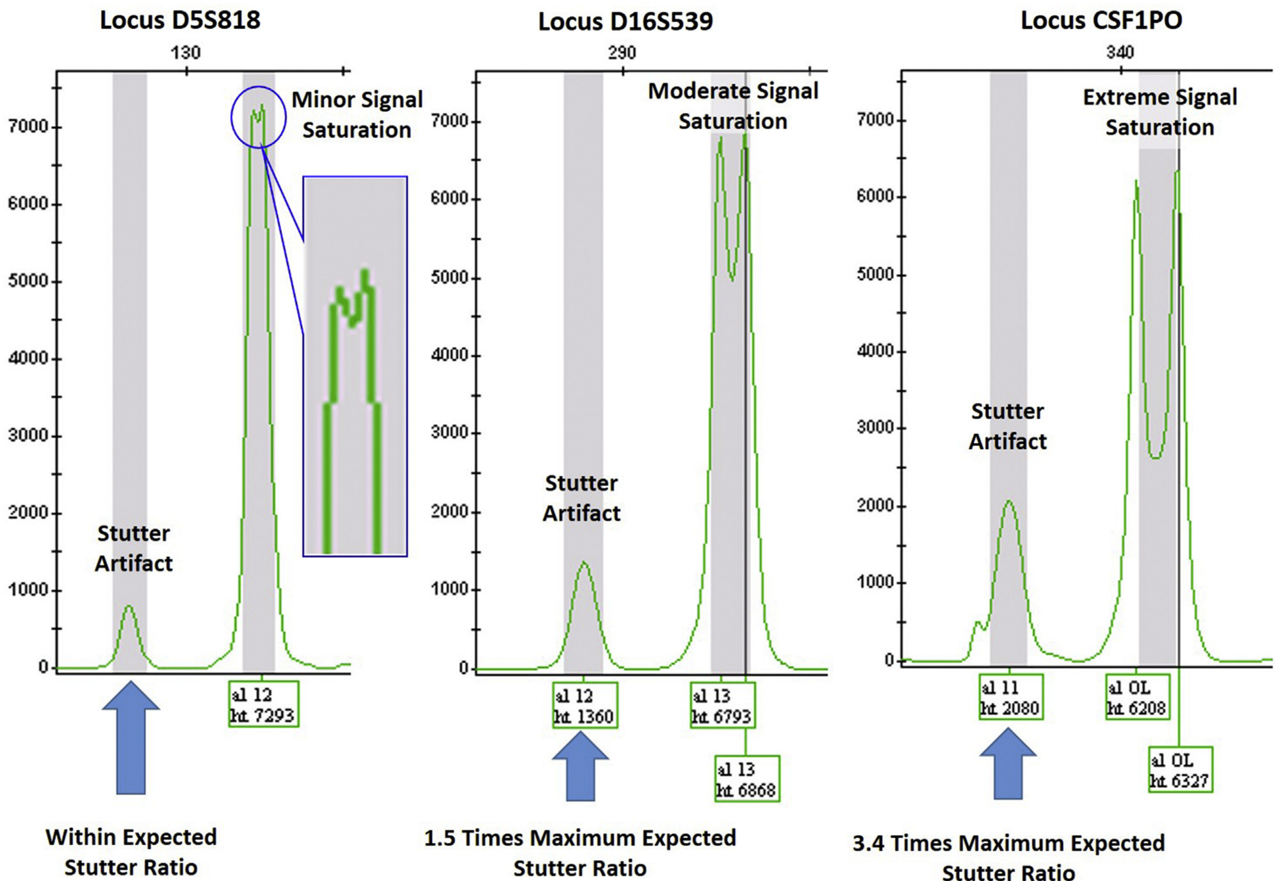


FIGURE 21.6 Relationship between signal saturation, off-scale peak morphology, and quantitative aspects of profile interpretation (e.g., ratio of stutter peak to allelic peak). As signal saturation increases, the allelic peak develops a split morphology. The greater the saturation, the deeper/more pronounced the mid-peak split. Left: At low levels of signal saturation, there is not a discernable quantitative impact on the expected ratio of the stutter peak to the allelic peak. Middle and Right: With greater signal saturation, the apparent height of the allelic peak decreases, resulting in stutter peaks that exceed the expected ratio of the stutter peak to the allelic peak resulting in a "false allele" being detected.

purification steps are warranted prior to amplification. Multiple qPCR kits are commercially available for the simultaneous quantitation of total human as well as male-specific DNA. The Quantifiler® Duo kit by Thermo Fisher Scientific (Barbisin et al., 2009), the Investigator Quantiplex HYres Kit by Qiagen (Frégeau and Laurin, 2015), and the Plexor® HY kit by Promega (Bulander and Rolf, 2009), for example, target a human-specific gene as well as a male-specific Y chromosome marker. This makes it possible to accurately quantify the amount of total human DNA and the amount of male DNA in a sample. This information is critical for choosing between subsequent DNA profiling using commercial kits based on autosomal STR versus Y-STR genetic markers. Autosomal STRs are best suited to samples with less than a 10- to 20-fold excess of female DNA compared to male DNA, while Y-STR markers can be used to detect a clear male DNA profile even in the presence of a 1000-fold excess of female DNA compared to male DNA. In the absence of information on the relative quantities of total human and male-specific DNA, a standard reliance on autosomal STR kits may fail to reveal the presence of a potentially probative male DNA profile (Horsman et al., 2006). The newest generation of qPCR kits, exemplified by the Quantifiler® Trio kit from Thermo Fisher Scientific, have incorporated both large autosomal and small autosomal amplification targets. This not only provides for some redundancy in DNA quantity estimates, but it also provides for a qualitative assessment of DNA integrity. The degradation of DNA will tend to preferentially impact larger amplicons over smaller amplicons. The ratio of the DNA quantity determined for the small versus large autosomal target, therefore, serves as a DNA degradation index. This is a general indicator of whether large DNA fragments may perform more poorly relative to small DNA fragments in subsequent STR reactions, and as such it helps to further inform workflow decisions (Vernarecci et al., 2015).

21.5 CAPILLARY ELECTROPHORESIS AND DATA INTERPRETATION

After amplification, various size fragments of DNA must be separated for analysis. As DNA molecules are negatively charged due to the phosphate groups on the nucleotides, they will migrate in an electric field, enabling electrophoresis to be used for separation. Capillary electrophoresis (CE) remains the most commonly used method for separating and detecting DNA for forensic purposes (Fig. 21.7). A hollow silica capillary spans two electrodes connected to a power supply. Buffer vials reside at each end of the capillary, which is filled with a linear polymer solution. Amplicons are treated with formamide, denaturing the DNA and allowing for a better resolution. Electrokinetic injection of amplicons directly onto the capillary occurs based on the negative charge of the DNA molecules. As the DNA migrates through the capillary toward the anode, it passes through the polymer sieving medium, allowing for the separation of DNA molecules based on size. Fluorescent dyes covalently bound to the 5′ end of one primer for each primer set per locus were incorporated into amplified product during PCR. Multiple color dyes are used for multiplexed reactions to allow for the analysis of loci with overlapping size ranges. As the migrating DNA fragments move through the capillary, they pass a clear detection window where a laser is used to excite the dye labels on the DNA fragments. Fluorescence is detected and measured, producing an eletropherogram. An internal size standard, or a constant set of known size fragments of DNA, is added to each sample to facilitate allele designation and the comparison of results between runs. Additionally, an allelic ladder that contains the most common alleles in the general population for each locus is used for accurate genotype determination (Butler et al., 2004).

21.5.1 Interpretation of Results

The interpretation of DNA profiling data is a matter of professional training and scientific expertise, firmly grounded in rigorous validation studies, peer-reviewed reports from the professional literature, and the fundamental principles of science. Accurate interpretations require the ability to discern an amplicon signal from instrument noise, to discriminate between signal artifacts and true contributor alleles, and to account for stochastic amplification effects (e.g., allele dropout and heterozygote imbalance) (Fig. 21.8). During PCR and CE, a number of artifacts may be created and detected that can interfere with the interpretation of resulting electropherograms. As previously mentioned, stutter artifacts can arise from strand slippage during PCR, resulting in artifact peaks one or more repeat units less than the true allele (Hauge and Litt, 1993). Stutter can also produce artifact peaks that are one repeat unit larger than the true allele although these are more rarely observed. The presence of stutter products can complicate DNA profile interpretation, especially in mixtures where stutter peaks may be indistinguishable from peaks attributed to a minor contributor. However, stutter formation follows reproducible percentages for each of the core STR loci used. In general, stutter percentages increase with the largest alleles within a locus. Overall, however, most stutter

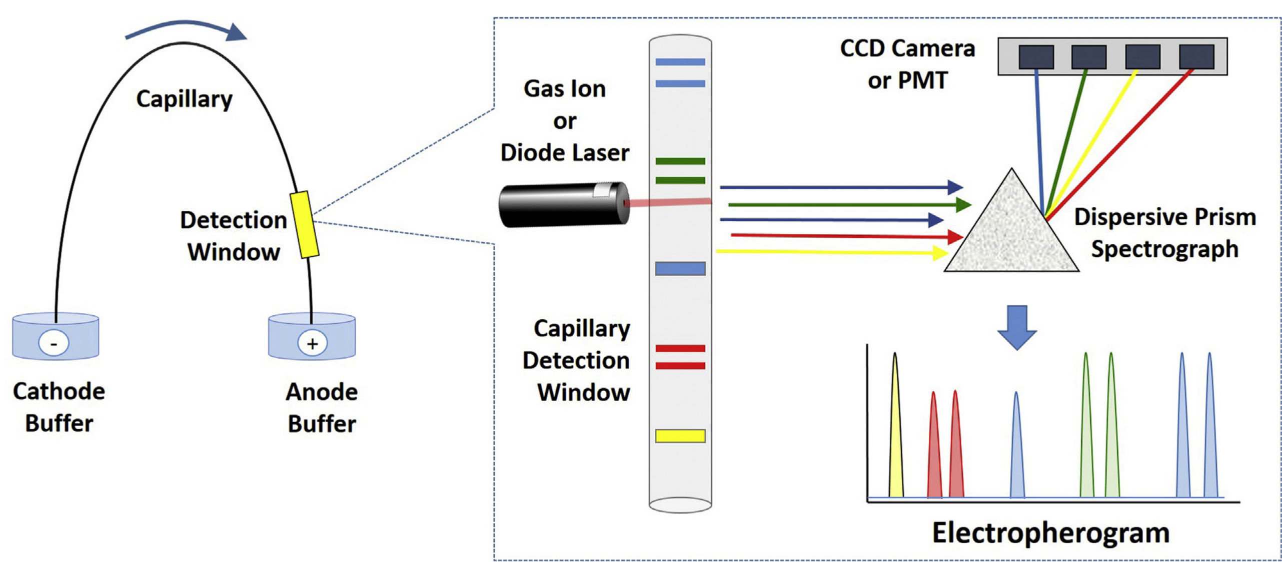


FIGURE 21.7 Schematic of the capillary electrophoretic separation and detection process for short tandem repeat (STR) alleles. Fluorescently labeled STR amplicons are size-fractionated in a capillary containing linear polyacrylamide as a sieving matrix. As alleles pass a clear detection window near the end of the capillary, a laser is used to induce florescence of the labeled amplicons. The fluorescent signal is separated by a dispersive prism spectrograph and quantified by a charge coupled device camera or photomultiplier tube. Signal processing software then produces an electropherogram to graphically represent the signal intensity data for interpretation by the analyst. *CCD*, charge coupled device; *PMT*, photomultiplier tube.

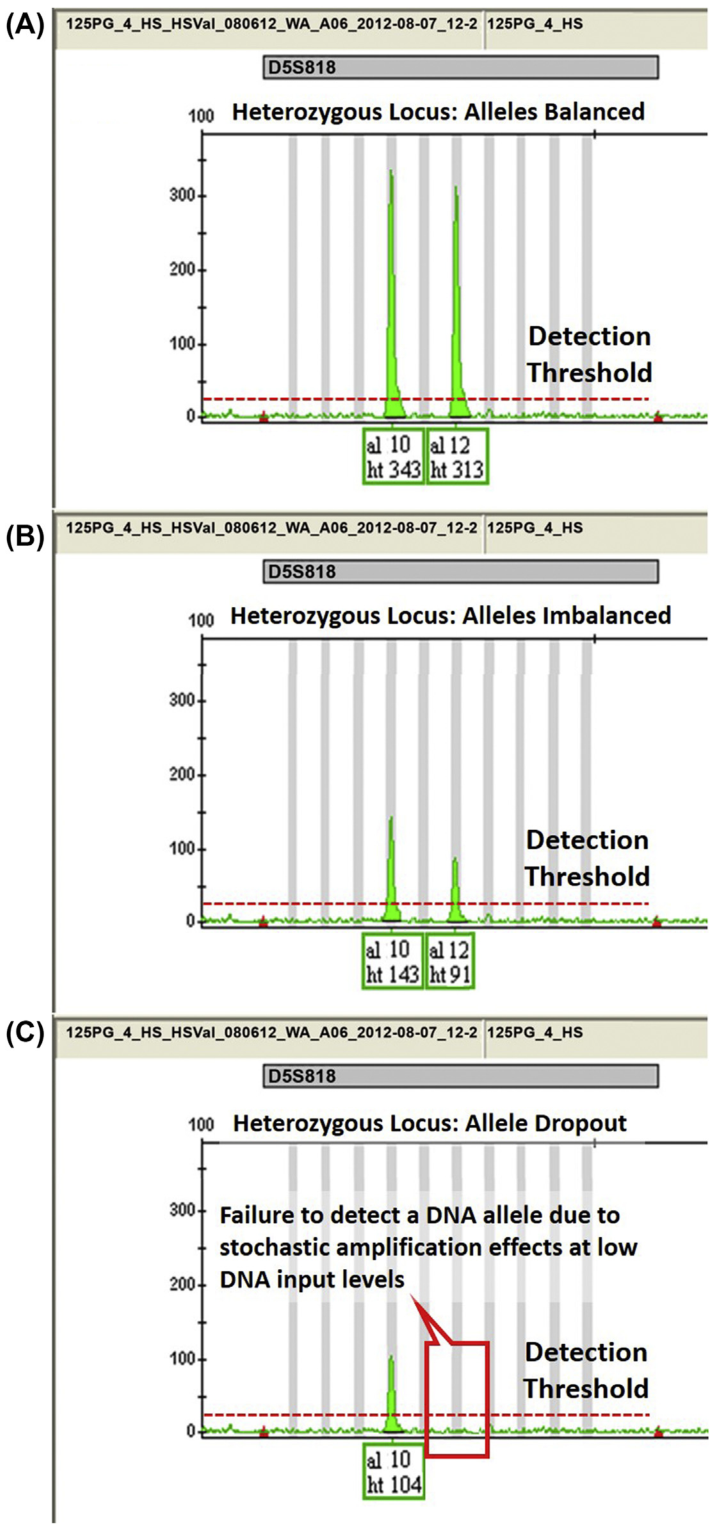


FIGURE 21.8 Illustrative example of the impact of stochastic amplification at low DNA input quantities on heterozygote allele detection accuracy. (A) An optimal DNA input typically allows for the accurate detection of both expected alleles in a DNA profile characterized by a good signal strength and peak height balance. (B) As the input DNA quantity becomes limiting, the signal strength decreases, and a greater variation in the peak height balance is observed. (C) At low DNA input quantities, stochastic amplification may result in a failure to detect an expected allele (i.e., allele dropout). This loss of genetic data increases the potential for erroneous DNA profile interpretations.

products have a peak height that is 15% or less of the height of the associated parent/source allele. Another artifact is an incomplete terminal nucleotide addition, which is characterized by the presence of an artifactual peak one base pair shorter than the expected allele. This results when the DNA polymerase fails to add an extra nucleotide (usually an adenine) to the 3′ end of a PCR product. This may be caused by the presence of excessive amounts of template DNA in the PCR reaction and/or suboptimal PCR conditions, especially too short of a final extension period during PCR. Commercial STR amplification kits recommend PCR conditions that favor the complete adenylation of all products. The presence of too much template DNA can also overwhelm the spectral calibration file thereby creating pull-up artifacts in a sample (Frank et al., 2001). This is when one color dye on the electropherogram is pulled up from a true allelic peak in another color due to overlapping spectra. This produces an artifact with the shape of an allelic peak. Voltage spikes and urea crystals in the capillary can also lead to nonreproducible artifacts known as spikes in an electropherogram (Butler, 2005). These appear as sharp peaks in all colors of the electropherogram with consistent peak heights. The predefined criteria for each of these artifacts based on observations from validation studies helps to eliminate them from profiles, leading to accurate genotypic interpretation.

Forensic laboratories employ quality-assurance criteria for the evaluation of DNA profiling results based on control data collected as part of the laboratory's DNA testing activities. Commonly employed controls include reagent blank controls, DNA extraction controls, positive amplification controls, and negative amplification controls.

The reagent blank control (e.g., a control sample processed in parallel with the forensic casework sample but to which no DNA source material was added) serves to demonstrate that the DNA extraction and processing reagents do not contain human DNA. The DNA extraction control (e.g., a blood or saliva sample having a previously determined DNA profile) is utilized to demonstrate that the DNA extraction process successfully yielded DNA having the expected profile. The positive amplification control (e.g., 007 DNA, 2800M DNA, or a similarly recognized DNA standard) serves to demonstrate the integrity of the PCR amplification process. Finally, the negative amplification control (e.g., water or dilute Tris EDTA added to a PCR reaction in place of a DNA template) serves to demonstrate that the DNA typing kit itself does not contain detectable levels of human DNA.

Prior to making any comparison between the human DNA profile from an item of evidence (i.e., the forensic profile) and a known reference profile from an alleged victim or suspect, the forensic profile should be carefully evaluated without a priori assumptions to determine if the sample is a single-source or a mixed-contributor profile. If a mixed-source profile is observed, a determination of the minimum number of contributors must be made. These initial determinations are critical to the accurate interpretation of the sample and the choice of the statistical approach that will be used to assess the scientific weight of the interpretation. Artifact and allele identification and the designation of loci that are suitable for use in statistical analyses should be completed before comparison to any known samples so as to avoid potential cognitive or confirmation bias on the part of the analyst.

Only after all of these initial assessments have been made should the forensic profile be compared to the reference profiles of suspects and alleged victims. This comparison generally results in one of three basic conclusions (SWGDAM, 2010). In a worst-case scenario, the comparison may be "Inconclusive" or "Uninterpretable" due to a paucity of genetic information or the presence of data that fails to meet the minimum quality standards for interpretation. No conclusions can or should be drawn regarding the potential source of DNA detected on such forensic samples. If the forensic DNA profile is of acceptable quality but is inconsistent with that of a given reference profile, the individual represented by the reference DNA profile is said to be "Excluded" as the potential source of the forensic DNA profile. When an individual's DNA profile is consistent with that from the item of evidence, it is said that the reference DNA profile "Cannot Be Excluded" as a potential source of the forensic DNA profile. In this last case, it is standard practice to assign a statistical significance/weight to the nonexclusion that reflects the rarity of the forensic profile among unrelated individuals in the general population.

21.6 STATISTICAL CALCULATIONS

21.6.1 Random Match Probability

Exclusions and inconclusive results warrant no further statistical analysis. However, a finding of "Cannot Be Excluded" requires that the significance associated with the potential match be calculated via statistical analysis (SWGDAM, 2010). With respect to DNA profiles based on autosomal STRs, the statistical measure of the rarity of a DNA profile of interest depends on the number of contributors to the profile and whether there is a need to mathematically model the probability of such stochastic phenomena as allele dropout to account for inconsistencies between a reference profile and a forensic profile. In general, an RMP statistic is appropriate for good quality single contributor forensic profiles. The RMP statistic indicates how common/rare a DNA profile is in the general population of unrelated individuals (Bille et al., 2013). The RMP is calculated using the multiplicative rule of probability to combine the

individual genotype frequencies determined for each analyzed locus. In accordance with the Hardy-Weinberg formula, heterozygote frequency (f) is calculated for alleles "p" and "q" using the equation $f = 2pq$. Homozygote frequency (f) is calculated for allele "p" using the equation $f = p^2 + p(1-p)\theta$ where $\theta = 0.01$ for most general populations and $\theta = 0.03$ for smaller and more reproductively isolated populations (e.g., Native Americans residing within the boundaries of a reservation) (Board, 2000; Council, 1996). It is important to point out that the standard RMP does not indicate how common/rare a DNA profile is among persons who are biologically related; such would require the use of modified formulas (Puch-Solis et al., 2012). This statistic is also not the probability that a given individual is the true source of the DNA in a specific sample.

21.6.2 Combined Probability of Inclusion/Exclusion

A combined probability of inclusion/exclusion (CPI/E) statistic is suitable for use with high-quality mixed DNA profiles (i.e., a forensic profile containing DNA from two or more individuals) where, after accounting for alleles that are potentially shared by different contributors, there is no need to model potential stochastic effects. The CPI/E indicates the probability that a randomly selected unrelated individual from the general population will be included/excluded as a potential contributor to a mixed-forensic DNA profile. This statistic is also based on the multiplicative rule of probability whereby a probability of inclusion (PI) value is calculated for each locus using the formula $PI = (f \text{ allele } 1 + f \text{ allele } 2 + \ldots + f \text{ allele } n)^2$ where "f" = the allelic frequency. The PI values across all loci within a profile are then multiplied together to obtain the CPI. The combined probability of exclusion (CPE) is then readily calculated from the CPI using the formula $CPE = 1-CPI$. It is important to point out that this statistic is not the probability that a given individual is an actual contributor to a mixture (Zhong et al., 1999).

21.6.3 Likelihood Ratio

The CPI statistic has also been widely criticized as being overly inclusive (i.e., erring on the side of false inclusion), not making full use of all available data, and not being applicable to forensic profiles subject to stochastic amplification phenomena (Buckleton et al., 2006). As a result, the use of the CPI/E statistic has decreased. In its place, the likelihood ratio (LR) statistic is being employed to determine the statistical weight of a "Cannot be Excluded" conclusion in regard to DNA mixtures. The LR statistic compares the relative support for two competing and mutually exclusive hypotheses under a specific set of a priori assumptions. LR values can range from 0 to ∞ where values greater than 1.0 provide support for one hypothesis under its associated assumptions, and values less than 1.0 provide support for the alternate hypothesis under its set of associated assumptions. A key advantage of using LR statistics is the ability to mathematically model different assumptions with regard to the number of possible contributors: the inclusion of assumed versus suspected contributors and the probability of stochastic phenomena. As powerful as LR statistical tools are, it is important to point out that LR calculations are sensitive to changes in the underlying hypotheses and/or assumptions, and there is no requirement that either of the proposed hypotheses be true (Goos et al., 2002). Finally, an LR statistic is not the probability that the underlying assumptions of either hypothesis are necessarily true/false. Rather, an LR indicates the likelihood of a given outcome (i.e., the forensic DNA profile) if one starts with the assumption that a given individual (i.e., the suspect, the alleged victim, or an individual selected at random from the general population) is a contributor.

Statistical calculations are used to determine the weight of evidence for forensic DNA profiles in which the DNA profile from the evidence has been found to "match" the DNA profile of a victim or suspect. While an extensive discussion regarding the calculations and statistical tools used to evaluate the strength of DNA profiling results is beyond the scope of this chapter, it is worth indicating that there are numerous statistical approaches that can be used for this purpose, including the RMP, the LR, and the combined probability of exclusion/inclusion (CPE/CPI) as described earlier. Guidelines for statistical interpretation of DNA profiles have been issued by the National Research Council, the DNA Advisory Board, and the European DNA Profiling Group. Additionally, there are numerous studies comparing the performance of different statistical models for the interpretation of DNA profiles.

21.6.4 Counting Method

The general interpretation workflow for Y-STR results follows that of STR loci. This includes a quality check of the data, a determination of the minimum number of contributors, a comparison of forensic and reference profiles, and a determination of whether the reference haplotype is excluded or not excluded as the potential source of the forensic haplotype. When an individual male's haplotype cannot be excluded as a potential source of the haplotype developed from an item of evidence, it is standard practice to state that neither the individual male nor any of his paternally related male relatives nor an unknown number of males in the general population can be excluded. It is then customary to assign a statistical significance/weight to the nonexclusion, which reflects the rarity of the haplotype in the general population of males. It is here where Y-STR interpretation differs from autosomal STR

interpretation. Since Y-STRs are inherited paternally as a haplotype and not as locus-independent events, the combined product of individual allele frequencies cannot be applied, as it is with autosomal STR profiles (Roewer et al., 1996, 2000). Instead, the rarity of the haplotype can only be estimated by applying the counting method to a Y-STR population database, as is commonly employed in the United States. (Budowle et al., 2003). To assign a statistical weight to a Y-STR haplotype of interest, the haplotype is searched against a reference database of unrelated individuals, and the number of times it is observed is counted and reported against the total number of profiles comprising the pertinent database population(s) that was searched. Thus the power of discrimination that can be achieved using Y-STRs is limited by the size of the population database used.

The general interpretation workflow for mtDNA analyses initially follows that of Y-STR haplotypes as well. This includes a quality check of the data and a determination of whether the sequence data indicate a single source sample or a mixture. Unlike Y-STR haplotypes, however, both the forensic haplotype and the reference haplotype are described in terms of their sequence differences relative to a common standard (i.e., the Revised Cambridge Reference Sequence). This then makes it possible to determine whether the reference haplotype is excluded or not excluded as the potential source of the forensic haplotype. When an individual mtDNA haplotype cannot be excluded as a potential source of the haplotype developed from an item of evidence, it is accurate to state that neither the individual nor any of their maternally related relatives nor an unknown number of persons in the general population can be excluded. It is then customary to assign a statistical significance/weight to the nonexclusion, which reflects the rarity of the haplotype in the general population of humans. As with Y-STR interpretations, the uniparental inheritance of the full maternal haplotype requires the use of the counting method to estimate the rarity of a given mtDNA haplotype. To assign a statistical weight to a given mtDNA haplotype of interest, the haplotype is searched against a reference database of unrelated individuals such as the SWGDAM Mitochondrial DNA Population Database and the European DNA Profiling group's Mitochondrial DNA Population database. The number of times a haplotype is observed is counted and reported against the total number of haplotypes comprising the pertinent database population(s) that was searched. Thus the power of discrimination that can be achieved using mtDNA is also limited by the size of the population database used.

21.7 NEXT GENERATION OF FORENSIC DNA TECHNOLOGIES

As indicated earlier in this chapter, advances in DNA extraction methods, Human DNA quantification by qPCR, PCR-based DNA profiling, and automated CE have all helped to revolutionize the field of forensic molecular biology. Additionally, the commercial availability of autosomal STR, Y-STR, mtDNA, and SNP-based typing kits that employ a common set of genetic markers have facilitated greater interlaboratory consistency, the sharing of DNA profile data among laboratories around the world, and the creation of searchable databases of forensic evidence, criminal offender profiles, missing persons, and unidentified human remains. As much as has been achieved, it is important to note that improvements in existing technologies and entirely new technical capabilities are continuously emerging. Two rapidly expanding new technologies are fully integrated sample-to-answer DNA profiling systems and next-generation genome sequencing technologies.

21.7.1 Rapid DNA Profiling

The vast majority of DNA profiling conducted by public and private forensic casework labs employs manual DNA extraction, quantitation, and genetic marker amplification kits. While many of these kits have steps that are amenable to automation, the overall process can still be somewhat labor intensive and time-consuming. Improvements that have been made to reduce the turnaround time have focused on direct PCR amplification from biological material (Flores et al., 2014; Myers et al., 2012; Park et al., 2008; Wang et al., 2011). By eliminating the need for separate DNA extraction and quantitation assays prior to DNA profiling, direct amplification approaches make it possible for forensic laboratories to generate profiles more rapidly. Since forensic samples must be processed by a laboratory, and many labs process forensic evidences in batches, there is still some lag time between sample collection at the crime scene and the interpretation of a resulting profile at the laboratory. The role of the laboratory as the optimal venue for testing has traditionally been justified by the major instrumentation and skilled personnel requirements, the need for tight environmental controls, and contamination prevention measures, including the strict segregation of pre- and post-PCR amplification areas. Overall, these factors have made it difficult, if not impossible, to generate DNA profiles from evidentiary material in less than 1 day without resorting to overtime or multiple work shifts. Similarly, these factors have stood as major obstacles to the ability to conduct reliable DNA profiling at the crime scene, booking station, or other field environment. Efforts over the years to bring a fully functional DNA laboratory to the crime scene have been operationally successful, but they also represent an investment that is financially prohibitive for the majority of law enforcement agencies.

In spite of what have been considered to be insurmountable hurdles, a long-standing desire on the part of criminal investigators has been for the development of

more mobile technologies for forensic DNA profiling that could produce near immediate DNA profile data from crime scene evidence, preferably in a forward operational environment without the need for extensive scientific expertise. Driving this desire is the recognition that the ability to identify a perpetrator in a timely manner can often speed the location and arrest of a criminal, thereby preventing further criminal activity.

The development of what have been characterized as "sample-to-profile" and microfluidic Lab-on-a-Chip technologies offer an opportunity to finally achieve rapid DNA profiling in the field by fully integrating DNA extraction, purification, template input, STR amplification, amplicon size fractionation, allele detection, and genotype analysis (Lounsbury et al., 2013; Greenspoon et al., 2008; Goedecke et al., 2004). The technical feasibility and performance characteristics of sample-to-profile and microfabricated DNA analysis components have been reported in the literature (Schmalzing et al., 1997; Hurth et al., 2010; Tan et al., 2013). The first commercially available instrument platforms with reagent packages for forensic applications that attracted serious interest and adoption by forensic analysts were the RapidHIT 200 Human Identification System from IntegenX Inc. and the DNAscan Rapid DNA Analysis System from partners NetBio and General Electric Healthcare. Beginning with a swab of blood or buccal cells, these instruments produce interpretable DNA profiles in approximately 1.5 h with minimal hands-on effort and without the need for a highly skilled forensic DNA analyst.

The RapidHIT 200 is a fully integrated benchtop instrument that uses a combination of one sample cartridge with four chambers for analysis and one control cartridge that also has four chambers for analysis. In the case of the control cartridge, three chambers are reserved for a positive and negative control and an STR allelic ladder. The last chamber may be used for one sample. The cartridges are run simultaneously such that profiles can be developed for up to five samples at a time. After the cartridges are inserted into the instrument and the sample swabs are placed into the cartridge chambers, DNA is extracted and purified using the DNA IQ paramagnetic bead-based cell lysis system from the Promega Corporation. This ensures that a controlled amount of DNA is captured and released into the amplification chamber, thus eliminating the need to quantify the concentration of the extracted DNA. After extraction, the DNA from each sample is moved to individual amplification chambers where it is combined with multiplexed primers and an amplification master mix to develop a DNA profile. The RapidHIT® GlobalFiler® Express kit (Hennessy et al., 2014), for example, simultaneously amplifies 24 loci, which include the expanded set of CODIS core STR loci, SE33, two Y chromosome genetic markers, and the sex-typing marker Amelogenin. Following amplification, the PCR products and an internal size standard are ported to an 8-channel CE plate for amplicon fractionation and detection by a charge coupled device camera. To improve fractionation, the capillaries are on a temperature-regulated printed circuit board rather than in an air chamber. The collected DNA profile data are automatically analyzed using the GeneMarker® HID software from SoftGenetics, LLC. The system's small footprint, rugged optics, and ease of use allow for it be moved with relative ease and thus to be used outside of a standard laboratory environment (e.g., at a crime scene or booking station).

The DNAscan system has a different design than the RapidHIT 200 system, but it is also capable of generating DNA profiles from five forensic samples at a time in about 1.5 h. The core of the system is a single-use microfluidic cassette called a BioChipSet Cassette (BCSC) that comes preloaded with all of the required reagents for sample processing. The BCSC consists of a smart cartridge, a gel smart cartridge, an integrated biochip, and a separation/detection biochip. The smart cartridge has five separate purification units, each of which holds one sample swab for analysis. There is a single formamide storage reservoir, and each of the five purification units are linked to four liquid reservoirs that hold the extraction/purification reagents. The system employs a chaotrope-silica purification method that has been adapted for microfluidic DNA purification. Upon completion of the DNA extraction and purification step, the DNA is transferred to the integrated biochip, which contains the lyophilized reagents (i.e., PCR mix, internal lane standard, and allelic ladder) required for PCR amplification and analysis of targeted genetic markers. The DNAscan system uses standard Promega STR analysis chemistries. Upon completion of the PCR step, electrophoresis-ready DNA is transferred to the separation/detection biochip, which performs size fractionation of the STR amplicons by electrophoresis in six electrophoresis channels. The gel smart cartridge holds the linear polyacrylamide-based sieving matrix and buffer used for the electrophoretic separation. An optical interface at the detection window receives laser light to excite fluorescent dyes for detection by a photomultiplier tube. Data analysis is then performed by automated allele calling software that interprets the processed data in accordance with rules designed to replicate the analytical processes of a forensic analyst.

Both systems were originally released for use with reference samples primarily due to their need for relatively large quantities of input DNA and high analytical thresholds. When used with reference samples, the overall success rate and accuracy of the rapid DNA profiling systems is comparable to that seen using standard DNA profiling methods (Gangano et al., 2013; Hennessy et al., 2013). Accordingly, the rapid generation of STR profiles may make it possible in some cases to obtain vital DNA-based

investigative leads from some types of evidentiary material in urgent high-priority cases. These platforms, however, do not yet provide equivalence in detection sensitivity as compared to standard DNA profiling systems. For the present, this will limit their utility for the analysis of trace and other more challenging samples that are typically associated with crime scenes.

21.7.2 Next-Generation Sequencing

The success of DNA testing as applied to forensic casework has brought with it a significant increase in demand, both for the analysis of larger numbers of samples and for greater amounts of genetic information to be obtained from forensic type samples, including partially degraded samples, which can be particularly challenging. An emerging area of focus for potential forensic applications of existing cutting-edge molecular biological techniques to meet this demand has been the use of NGS technologies. The capabilities of NGS technologies are superior to traditional CE-based platforms. Since NGS runs can be adjusted in terms of coverage for sequence reads, the technique offers a great degree of flexibility. For example, NGS platforms provide the option of analyzing multiple marker types, such as STRs, SNPs, mtDNA, and Y chromosome markers, concurrently in a single run. Moreover, NGS technology's ability to generate the same amount of information in a fraction of the time that a CE platform would require has made it an indispensable tool for forensic genomic research.

NGS platforms have the ability to increase the amount of useful genetic information that can be obtained from forensic samples. Consider, for example, that CE platforms (as used in a forensic context) distinguish STR alleles based on amplicon size rather than direct sequence. As a result, alleles that are of identical length but which differ in sequence are not recognized. This results in a loss of potentially vital genetic information that could serve to more confidently include or exclude an individual as a potential contributor to the DNA profile developed from an item of evidence. Advances in NGS technology now produce read lengths sufficient for STR testing. Thus even when alleles are of identical length, sequence information can be used to more precisely facilitate the identification of distinct alleles. This, in turn, facilitates the accurate interpretation of complex genomic DNA mixtures (Yang et al., 2014). Digital readouts of the number of reads, or individual sequences, also provide quantitative information that can be applied to mixture deconvolution even for SNP and mtDNA markers.

NGS platforms are also increasing the amount of information that can be generated from partially degraded genomic DNA through the expanded use of SNP markers. The use of SNPs with CE-based technologies is quite limited because of the relatively small number of SNP loci that can be multiplexed into an individual assay. Given the biallelic nature of SNPs, this greatly reduces the achievable power of discrimination unless a greater number of SNP loci can be typed. However, using an NGS-based approach, thousands of SNP markers can be analyzed simultaneously, thereby dramatically increasing the amount of useful data that can be obtained from partially degraded samples (Daniel et al., 2015).

Several NGS platforms are available on the market, and several manufacturers have explored the development of forensic applications. Among these, the MiSeq FGx System by Illumina is the first fully validated sequencing system designed for forensic applications that target a subset of informative genetic markers. The MiSeq FGx instrument platform, library preparation kits, customized instrument control, and data analysis software have been developmentally validated in accordance with SWGDAM guidelines for forensic applications. It is specifically designed to support the reliable analysis of both routine and challenging forensic samples. Coupled with the ForenSeq DNA Prep kit, the MiSeq FGx platform multiplexes the targeted sequencing of multiple types of forensically relevant genetic markers. These include 63 STRs and 95 identity SNPs, 56 ancestry SNPs, and 22 phenotypic SNPs. NGS-based sequencing also provides new opportunities for the accurate quantitative deconvolution of DNA mixtures and the identification of intra-STR allele sequence variants that would be missed by CE-based platforms. The SNPs that are assayed by the ForenSeq DNA Prep Kit are less than 125 base pairs in length, which makes them ideal targets for the analysis of partially or highly degraded DNA. In fact, it has been shown that >90% of SNP markers were correctly typed using less than 100 pg of input DNA (Churchill et al., 2016). The sensitivity of this platform and low DNA input requirements make this system better equipped to handle the challenges facing the field of forensic biology today as compared to traditional CE-based approaches. Thermo Fisher Scientific has also marketed two forensically targeted panels: the HID Ion AmpliSeq Ancestry Panel and the Identity Panel. The Ancestry Panel targets 165 autosomal SNP markers while the Identity Panel targets 124 SNP markers, including 90 autosomal SNPs and 34 Y-SNPs.

Future NGS platforms with potential forensic utility are focused on single molecule real-time sequencing. Pacific Biosciences is leading the field with the Sequel System and Oxford Nanopore with the MinION, PromethION, and GridION systems. These platforms not only detect single molecules but also promise longer read lengths and the ability to directly detect epigenetic modifications, which in a forensic application can be used to genetically distinguish identical twins.

Overall, forensic geneticists worldwide are now looking to NGS platforms and applications as a reliable means of

overcoming the limitations of the classical CE-based approach to genomic analyses. NGS promises the field of forensics increasing genetic discrimination by increasing the total number of informative genetic markers that can be detected in a cost-effective manner.

21.8 CONCLUSIONS

Modern advances in the field of molecular biology have had a significant impact on the applied field of forensic science and forensic identification cases. With a multitude of available genetic markers and improved analysis platforms, even small and/or degraded samples can be genetically categorized and have the potential to provide meaningful information to criminal investigators. The use of STRs, Y chromosome markers, and mtDNA with CE platforms provides sufficient discrimination power for most applications and has become well established and routinely employed in the field. Most industrialized nations have now established or are in the process of establishing forensic DNA databases based on STR technology for aiding criminal investigations. There is rapidly growing interest in the potential utility of SNP markers for gaining phenotypic information and for the analysis of degraded samples. Similarly, NGS platforms are being developed to squeeze more genetic information from standard and challenging forensic casework samples. The simultaneous detection of autosomal STR markers, Y chromosome markers, mtDNA, and ancestry, identity, and phenotypic SNPs will provide critical genetic information to forensic investigations more readily and at a lower cost than was previously possible.

Oftentimes advances in forensic technologies have grown out of those of the medical diagnostics community, a fact that explains why forensic DNA testing laboratories often seem to be technologically lagging behind their clinical counterparts. At the same time, the fact that forensic technologies have roots in the medical molecular diagnostics community serves to provide a critical foundation of support for the admissibility of such technologies in the courts. This is because the judicial systems in the United States and around the world tend to be conservative. The US courts look for either indications of general acceptance or of the reliability and relevance of a new technology before admitting it in judicial proceedings. An error in a medical diagnostic test can certainly have dire consequences for a patient and may result in a civil lawsuit against the physician, the clinical testing facility, or the assay manufacturer. In the criminal justice context, however, an error in a forensic test can result in an innocent person being wrongfully convicted and deprived of his or her freedom or wrongfully executed. In a criminal context, an error in a forensic test may result in a failure to convict a guilty person who is then released to commit additional heinous crimes. The close interrelationship between the fields of forensics and molecular diagnostics therefore is an important part of helping to ensure the confidence of the general public in our institutions of justice.

REFERENCES

Allen, M., Engström, A.-S., Meyers, S., Handt, O., Saldeen, T., von Haeseler, A., Pääbo, S., Gyllensten, U., 1998. Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: sensitivity and matching probabilities. J. Forensic Sci. 43, 453−464.

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., Howell, N., 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat. Genet. 23, 147.

Ayub, Q., Mohyuddin, A., Qamar, R., Mazhar, K., Zerjal, T., Mehdi, S.Q., Tyler-Smith, C., 2000. Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information. Nucleic Acids Res. 28, e8.

Bacher, J., Hennes, L.F., Gu, T., Tereba, A., Micka, K.A., Sprecher, C.J., Lins, A.M., Amiott, E.A., Rabbach, D.R., Taylor, J.A., 1999. Pentanucleotide repeats: highly polymorphic genetic markers displaying minimal stutter artifact. In: Proceedings from the Ninth International Symposium on Human Identification.

Bacher, J., Schumm, J.W., 1998. Development of highly polymorphic pentanucleotide tandem repeat loci with low stutter. Profiles DNA 2, 3−6.

Barbisin, M., Fang, R., O'shea, C.E., Calandro, L.M., Furtado, M.R., Shewale, J.G., 2009. Developmental validation of the Quantifiler® Duo DNA Quantification kit for simultaneous quantification of total human and human male DNA and detection of PCR inhibitors in biological samples. J. Forensic Sci. 54, 305−319.

Bargagna, M., Pereira, M., 1967. A study of absorption-elution as a method of identification of Rhesus antigens in dried bloodstains. J. Forensic Sci. Soc. 7, 123−130.

Bille, T., Bright, J.A., Buckleton, J., 2013. Application of random match probability calculations to mixed STR profiles. J. Forensic Sci. 58, 474−485.

Board, D.A., 2000. Statistical and population genetics issues affecting the evaluation of the frequency of occurrence of DNA profiles calculated from pertinent population database (s). Forensic Sci. Commun. 2.

Bogenhagen, D., Clayton, D.A., 1974. The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. Quantitative isolation of mitochondrial deoxyribonucleic acid. J. Biol. Chem. 249, 7991−7995.

Brevnov, M.G., Pawar, H.S., Mundt, J., Calandro, L.M., Furtado, M.R., Shewale, J.G., 2009. Developmental validation of the PrepFiler™ Forensic DNA Extraction Kit for extraction of genomic DNA from biological samples. J. Forensic Sci. 54, 599−607.

Buckleton, J., Triggs, C., Champod, C., 2006. An extended likelihood ratio framework for interpreting evidence. Sci. Justice 46, 69−78.

Budowle, B., Moretti, T.R., Niezgoda, S.J., Brown, B.L., 1998. Codis and PCR-based Short Tandem Repeat Loci: Law Enforcement Tools. Second European Symposium on Human Identification. Promega Corporation, Madison, Wisconsin, pp. 73−88.

Budowle, B., Murch, R.S., 1986. Applications of isoelectric focusing in forensic serology. J. Forensic Sci. 31, 869−880.

Budowle, B., Sinha, S., Lee, H., Chakraborty, R., 2003. Utility of Y-Chromosome short tandem repeat haplotypes in forensic applications. Forensic Sci. Rev. 15, 153−164.

Bulander, N., Rolf, B., 2009. Comparison of the Plexor® HY system, Quantifiler® and Quantifiler Duo® kits using the Roche LightCycler 480 system and the ABI 7900 real time PCR instrument. Forensic Sci. Int. Genet. Suppl. Ser. 2, 104−105.

Butler, J.M., 2005. Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers. Academic Press.

Butler, J.M., Buel, E., Crivellente, F., McCord, B.R., 2004. Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. Electrophoresis 25, 1397−1412.

Churchill, J.D., Schmedes, S.E., King, J.L., Budowle, B., 2016. Evaluation of the Illumina® Beta version ForenSeq DNA Signature Prep Kit for use in genetic profiling. Forensic Sci. Int. Genet. 20, 20−29.

Clayton, T., Whitaker, J., Maguire, C., 1995. Identification of bodies from the scene of a mass disaster using DNA amplification of short tandem repeat (STR) loci. Forensic Sci. Int. 76, 7−15.

Collins, J.R., Stephens, R.M., Gold, B., Long, B., Dean, M., Burt, S.K., 2003. An exhaustive DNA micro-satellite map of the human genome using high performance computing. Genomics 82, 10−19.

Council, N.R., 1996. The Evaluation of Forensic DNA Evidence. The National Academies Press, Washington, DC.

Daniel, R., Santos, C., Phillips, C., Fondevila, M., van Oorschot, R.A., Carracedo, A., Lareu, M.V., McNevin, D., 2015. A SNaPshot of next generation sequencing for forensic SNP analysis. Forensic Sci. Int. Genet. 14, 50−60.

FBI, 2008. Quality assurance standards for forensic DNA testing laboratories. Forensic Sci. Commun. 10.

Edwards, A., Civitello, A., Hammond, H.A., Caskey, C.T., 1991. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am. J. Hum. Genet. 49, 746.

Ellegren, H., 2004. Microsatellites: simple sequences with complex evolution. Nat. Rev. Genet. 5, 435−445.

Ensenberger, M.G., Thompson, J., Hill, B., Homick, K., Kearney, V., Mayntz-Press, K.A., Mazur, P., McGuckian, A., Myers, J., Raley, K., 2010. Developmental validation of the PowerPlex® 16 HS System: an improved 16-locus fluorescent STR multiplex. Forensic Sci. Int. Genet. 4, 257−264.

Flores, S., Sun, J., King, J., Budowle, B., 2014. Internal validation of the GlobalFiler™ Express PCR Amplification Kit for the direct amplification of reference DNA samples on a high-throughput automated workflow. Forensic Sci. Int. Genet. 10, 33−39.

Frank, W.E., Llewellyn, B.E., Fish, P.A., Riech, A.K., Marcacci, T.L., Gandor, D.W., Parker, D., Carter, R.R., Thibault, S.M., 2001. Validation of the AmpFℓSTR™ profiler Plus PCR amplification kit for use in forensic casework. J. Forensic Sci. 46, 642−646.

Frégeau, C.J., Laurin, N., 2015. The Qiagen Investigator® Quantiplex HYres as an alternative kit for DNA quantification. Forensic Sci. Int. Genet. 16, 148−162.

Gangano, S., Elliott, K., Anoruo, K., Gass, J., Buscaino, J., Jovanovich, S., Harris, D., 2013. DNA investigative lead development from blood and saliva samples in less than two hours using the RapidHIT™ Human DNA Identification System. Forensic Sci. Int. Genet. Suppl. Ser. 4, e43−e44.

Ge, J., Eisenberg, A., Budowle, B., 2012. Developing criteria and data to determine best options for expanding the core CODIS loci. Investig. Genet. 3, 1.

Gill, P., Brenner, C., Brinkmann, B., Budowle, B., Carracedo, A., Jobling, M.A., De Knijff, P., Kayser, M., Krawczak, M., Mayr, W., 2001. DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome short tandem repeats. Leg. Med. 3, 252−257.

Gill, P., Ivanov, P.L., Kimpton, C., Piercy, R., Benson, N., Tully, G., Evett, I., Hagelberg, E., Sullivan, K., 1994. Identification of the remains of the Romanov family by DNA analysis. Nat. Genet. 6, 130−135.

Gill, P., Jeffreys, A.J., Werrett, D.J., 1985. Forensic application of DNA "fingerprints". Nature 318, 577−579.

Goedecke, N., McKenna, B., El-Difrawy, S., Carey, L., Matsudaira, P., Ehrlich, D., 2004. A high-performance multilane microdevice system designed for the DNA forensics laboratory. Electrophoresis 25, 1678−1686.

Goos, L., Silverman, I., Rose, D., Newman, J., 2002. The influence of probabilistic statements on the evaluation of the significance of a DNA match. Can. Soc. Forensic Sci. J. 35, 77−90.

Greenberg, B.D., Newbold, J.E., Sugino, A., 1983. Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. Gene 21, 33−49.

Greenspoon, S.A., Yeung, S.H., Johnson, K.R., Chu, W.K., Rhee, H.N., McGuckian, A.B., Crouse, C.A., Chiesl, T.N., Barron, A.E., Scherer, J.R., 2008. A forensic laboratory tests the Berkeley microfabricated capillary array electrophoresis device. J. Forensic Sci. 53, 828−837.

Hagelberg, E., Gray, I.C., Jeffreys, A.J., 1991. Identification of the skeletal remains of a murder victim by DNA analysis. Nature 352, 427−429.

Hagins, A., Mortimer, C., Shaler, R., 1978. MN determinations in bloodstains—selective destruction of cross-reacting activity. J. Forensic Sci. 23, 570−576.

Hauge, X., Litt, M., 1993. A study of the origin of "shadow bands" seen when typing dinucleotide repeat polymorphisms by the PCR. Hum. Mol. Genet. 2, 411−415.

Hennessy, L.K., Franklin, H., Li, Y., Buscaino, J., Chear, K., Gass, J., Mehendale, N., Williams, S., Jovanovich, S., Harris, D., 2013. Developmental validation studies on the RapidHIT™ human DNA identification system. Forensic Sci. Int. Genet. Suppl. Ser. 4, e7−e8.

Hennessy, L.K., Mehendale, N., Chear, K., Jovanovich, S., Williams, S., Park, C., Gangano, S., 2014. Developmental validation of the GlobalFiler® express kit, a 24-marker STR assay, on the RapidHIT® System. Forensic Sci. Int. Genet. 13, 247−258.

Holland, M.M., Fisher, D.L., Mitchell, L.G., Rodriquez, W., Canik, J.J., Merril, C.R., Weedn, V.W., 1993. Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam War. J. Forensic Sci. 38, 542−553.

Horsman, K.M., Hickey, J.A., Cotton, R.W., Landers, J.P., Maddox, L.O., 2006. Development of a human-specific real-time PCR assay for the simultaneous quantitation of total genomic and male DNA. J. Forensic Sci. 51, 758−765.

Hurth, C., Smith, S.D., Nordquist, A.R., Lenigk, R., Duane, B., Nguyen, D., Surve, A., Hopwood, A.J., Estes, M.D., Yang, J., 2010. An automated instrument for human STR identification: design, characterization, and experimental validation. Electrophoresis 31, 3510−3517.

Ivanov, P.L., Wadhams, M.J., Roby, R.K., Holland, M.M., Weedn, V.W., Parsons, T.J., 1996. Mitochondrial DNA sequence heteroplasmy in the

Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. Nat. Genet. 12, 417−420.

Jeffreys, A., 1985. Individual-specific “fingerprints” of human DNA. Nature 316, 4.

Jeffreys, A., Wilson, V., Thein, S., 1992. Hypervariable “minisatellite” regions in human DNA. 1985. Biotechnology 24, 467. Read Mass.

Jeffreys, A.J., Brookfield, J.F., Semeonoff, R., 1985. Positive identification of an immigration test-case using human DNA fingerprints. Nature 317, 818−819.

Karafet, T.M., Mendez, F.L., Meilerman, M.B., Underhill, P.A., Zegura, S.L., Hammer, M.F., 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res. 18, 830−838.

Kayser, M., 2003. The human Y-Chromosome-introduction into genetics and applications. Forensic Sci. Rev. 15, 77−89.

Köchl, S., Niederstätter, H., Parson, W., 2005. DNA extraction and quantitation of forensic samples using the phenol-chloroform method and real-time PCR. Forensic DNA Typing Protoc. 13−29.

Koons, B.W., Sobieralski, C.A., Comey, C.T., Baechtel, E.S., Smerick, J.B., Presley, K.W., Stanley, D.M., 1994. DNA extraction strategies for amplified fragment length polymorphism analysis. J. Forensic Sci. 39, 1254−1269.

Krnajski, Z., Geering, S., Steadman, S., 2007. Performance verification of the Maxwell 16 Instrument and DNA IQ Reference Sample Kit for automated DNA extraction of known reference samples. Forensic Sci. Med. Pathol. 3, 264−269.

Li, R., 2015. Forensic Biology. CRC Press, p. 567.

Lounsbury, J.A., Karlsson, A., Miranian, D.C., Cronk, S.M., Nelson, D.A., Li, J., Haverstick, D.M., Kinnon, P., Saul, D.J., Landers, J.P., 2013. From sample to PCR product in under 45 minutes: a polymeric integrated microdevice for clinical and forensic DNA analysis. Lab Chip 13, 1384−1393.

Myers, B.A., King, J.L., Budowle, B., 2012. Evaluation and comparative analysis of direct amplification of STRs using PowerPlex® 18D and Identifiler® Direct systems. Forensic Sci. Int. Genet. 6, 640−645.

Pakstis, A.J., Speed, W.C., Fang, R., Hyland, F.C., Furtado, M.R., Kidd, J.R., Kidd, K.K., 2010. SNPs for a universal individual identification panel. Hum. Genet. 127, 315−324.

Park, S.J., Kim, J.Y., Yang, Y.G., Lee, S.H., 2008. Direct STR amplification from whole blood and blood-or saliva-spotted FTA® without DNA purification. J. Forensic Sci. 53, 335−341.

Phillips, C., Fang, R., Ballard, D., Fondevila, M., Harrison, C., Hyland, F., Musgrave-Brown, E., Proff, C., Ramos-Luis, E., Sobrino, B., 2007. Evaluation of the Genplex SNP typing system and a 49plex forensic marker panel. Forensic Sci. Int. Genet. 1, 180−185.

Prinz, M., Carracedo, A., Mayr, W., Morling, N., Parsons, T., Sajantila, A., Scheithauer, R., Schmitter, H., Schneider, P.M., 2007. DNA Commission of the International Society for Forensic Genetics (ISFG): recommendations regarding the role of forensic genetics for disaster victim identification (DVI). Forensic Sci. Int. Genet. 1, 3−12.

Puch-Solis, R., Roberts, P., Pope, S., Aitken, C., 2012. Assessing the Probative Value of DNA Evidence: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses. RSS.

Robin, E.D., Wong, R., 1988. Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. J. Cell. Physiol. 136, 507−513.

Roewer, L., Kayser, M., De Knijff, P., Anslinger, K., Betz, A., Caglia, A., Corach, D., Füredi, S., Henke, L., Hidding, M., 2000. A new method for the evaluation of matches in non-recombining genomes: application to Y-chromosomal short tandem repeat (STR) haplotypes in European males. Forensic Sci. Int. 114, 31−43.

Roewer, L., Kayser, M., Dieltjes, P., Nagy, M., Bakker, E., Krawczak, M., De Knijff, P., 1996. Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related human populations. Hum. Mol. Genet. 5, 1029−1033.

SWGDAM, 2010. Interpretation Guidelines for Autosomal STR Typing [Online] Available. https://www.fbi.gov/about-us/lab/biometric-analysis/codis/swgdam.pdf.

Satoh, M., Kuroiwa, T., 1991. Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. Exp. Cell Res. 196, 137−140.

Scarpetta, M.A., Drayton, M.L., Turek, S.A., Greenspoon, S.A., 1998. QIAamp spin columns as a method of DNA isolation for forensic casework. J. Forensic Sci. 43, 1024−1030.

Schmalzing, D., Koutny, L., Adourian, A., Belgrader, P., Matsudaira, P., Ehrlich, D., 1997. DNA typing in thirty seconds with a microfabricated device. Proc. Natl. Acad. Sci. U. S. A. 94, 10273−10278.

Schneider, P.M., Meuser, S., Waiyawuth, W., Seo, Y., Rittner, C., 1998. Tandem repeat structure of the duplicated Y-chromosomal STR locus DYS385 and frequency studies in the German and three Asian populations. Forensic Sci. Int. 97, 61−70.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423, 825−837.

Springer, G.F., Desai, P.R., 1975. Human blood-group MN and precursor specificities: structural and biological aspects. Carbohydr. Res. 40, 183−192.

Subramanian, S., Mishra, R.K., Singh, L., 2003. Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. Genome Biol. 4, 1.

Tan, E., Turingan, R.S., Hogan, C., Vasantgadkar, S., Palombo, L., Schumm, J.W., Selden, R.F., 2013. Fully integrated, fully automated generation of short tandem repeat profiles. Investig. Genet. 4, 1.

Taylor, R.W., Turnbull, D.M., 2005. Mitochondrial DNA mutations in human disease. Nat. Rev. Genet. 6, 389−402.

Tumpenny, P., Ellard, S., 2005. Emery’s Elements of Medical Genetics. Twelfth ed. Churchill Livingstone, Elsevier, London, p. 416.

Vernarecci, S., Ottaviani, E., Agostino, A., Mei, E., Calandro, L., Montagna, P., 2015. Quantifiler® Trio Kit and forensic samples management: a matter of degradation. Forensic Sci. Int. Genet. 16, 77−85.

Walsh, P.S., Fildes, N.J., Reynolds, R., 1996. Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. Nucleic Acids Res. 24, 2807−2812.

Walsh, P.S., Metzger, D.A., Higuchi, R., 1991. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques 10, 506−513.

Wang, D.Y., Chang, C.W., Lagacé, R.E., Oldroyd, N.J., Hennessy, L.K., 2011. Development and validation of the AmpFℓSTR® Identifiler® direct PCR Amplification Kit: a multiplex assay for the direct amplification of single-source samples. J. Forensic Sci. 56, 835−845.

Wyman, A.R., White, R., 1980. A highly polymorphic locus in human DNA. Proc. Natl. Acad. Sci. U. S. A. 77, 6754–6758.

Yamamoto, F.-I., Clausen, H., White, T., Marken, J., Hakomori, S.-I., 1990. Molecular genetic basis of the histo-blood group ABO system. Nature 345, 229–233.

Yang, Y., Xie, B., Yan, J., 2014. Application of next-generation sequencing technology in forensic science. Genom. Proteom. Bioinforma. 12, 190–197.

Yoshida, K., Sekiguchi, K., Mizuno, N., Kasai, K., Sakai, I., Sato, H., Seta, S., 1995. 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.

Zhong, Y., Budowle, B., Center, F., 1999. The utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems. Electrophoresis 20, 1682–1696.

Chapter 22

New Perspectives in Mass Disaster Victim Identification Assisted by DNA Typing and Forensic Genomics

D. Corach
Universidad de Buenos Aires, Buenos Aires, Argentina

22.1 INTRODUCTION

At the beginning of the 20th century, Karl Landsteiner (1901) identified the ABO blood group, the first monogenic serological polymorphism employed for preventing transfusion-associated disease and for kinship rejections due to its Mendelian transmission. Over 5 decades elapsed between the detection of this genetic marker and the unveiling of the DNA structure (Watson and Crick, 1953) and over 80 years from the detection of human DNA polymorphisms (Wyman and White, 1980).

Hence, for over 80 years, blood groups and other serological polymorphisms remained as the molecular diagnostic tools for inferring biological kinship. Their reduced discriminatory power and the cell integrity required for the analysis limited the potential applicability for investigating forensic evidentiary material or decomposed corpses or remains.

After the serological genetic markers were discovered, achieving the state of the art in the forensic field in German-speaking countries, mainly Austria and Germany, a scientific society named the "Gesellschaft für forensische Blutgruppenkunde" (Society for Forensic Hemogenetics) was founded in 1968. Due to the growing interest in serological polymorphisms and their use in forensic investigations, this society became an international forum in 1989, receiving the status of an international society ("International Society for Forensic Haemogenetics").

In the mid-1980s, two discoveries and a technical development converged, and a few years later they combined synergically. The development of the polymerase chain reaction (PCR) by Mullis et al. (1986 and the discovery of hypervariable minisatellite sequences in human DNA by Jeffreys et al. (1985) and microsatellite DNA markers by Epplen et al. (1982), Panicker and Singh (1994), Tautz and Renz (1984), and Weber and Mayr (1989) strongly impacted the forensic community by introducing approaches that improved the usefulness of molecular polymorphisms in forensic sciences. These new approaches substantially modified the analytical framework. In 1991, as a result of the transition from serological to DNA markers, the International Society of Forensic Hemogenetics was renamed, receiving the denomination of "International Society for Forensic Genetics." This society is the leading international forum in the field.

A major impact was brought forth in the field of human identification in cases involving living or deceased people, since even a certain degree of DNA degradation was likely to be neglected by the analysis of microsatellites. Microsatellites are small polymorphic regions, also termed short tandem repeats (STRs), which are widely distributed throughout the genome. Their applicability and use in forensic identification was demonstrated by Edwards et al. (1991,1992), becoming the gold standard technique in the field of human identification.

The availability of this new analytical approach efficiently contributed to the development of strategies that allowed for the identification of decomposed corpses and fragmentary human remains, often "by-products" of mass disasters, by comparing their genetic profiles with those of their putative relatives. This chapter will briefly summarize the evolution of the DNA-based methods used for identifying victims and human remains from mass disasters and

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00022-5
Copyright © 2017 Elsevier Ltd. All rights reserved.

discuss the new forensic genomic tools that will further optimize forensic identification procedures. The new approaches offer, in addition to identification tools, other marker sets that might help in determining the geographical/ethnic affinities of the victims as well as providing information about phenotypic traits from fragmented corpses, refining the identification capabilities introduced within the DNA next-generation sequencing (NGS) approach.

22.2 CLASSIFICATION OF MASS FATALITIES AND DIVERSE SCENARIOS FOR HUMAN REMAINS RETRIEVAL

Hazards and mass disasters are extremely complex scenarios whose origins might be mainly determined by two major forces: natural and man-induced catastrophes. Some authors such as Wisner (2004) argue that all disasters are man-induced, since human action can prevent their development into a mass disaster before the strike of the hazard. Human failure in developing adequate disaster management preventive policies might result in human responsibility in most of them. Different kinds of hazards, which indicate the acting forces and their impact on the environment and human beings, are summarized in Table 22.1. It is important to keep in mind that some hazards are far from simple in terms of their effects, as they might produce a chain reaction with extremely severe effects.

Since the beginning of 2010, a number of catastrophic events have occurred worldwide, including natural disasters, airplane accidents, and terrorist attacks. These events are summarized in Fig. 22.1, and the underling information was obtained from journal reports, as in the case of Fig. 22.2.

Concerning natural disasters, in addition to the earthquake that severely affected Haiti on January 12, 2010, killing over 160,000 people, other natural disasters left a death toll of more than 1500 people. In the same period of time, commercial airplane crashes caused over 2700 fatalities and terrorist attacks claimed more than 20,000 victims, as depicted in Fig. 22.2. Many of the fatalities required DNA-assisted body fragment reconstruction followed by identification aided by DNA typing approaches.

Many factors may cause these catastrophes and, as a result, developing countries are directly affected by their effects. In fact, in these countries, human vulnerability is increased due to the lack of appropriate emergency management and planning. Accordingly, the unexpectedness of the event may intensify the adverse effects of the hazard, which, in turn, may result in extremely serious human, structural, and financial losses. The deadly earthquake in Haiti left behind a death toll of over 160,000 people. However, generalizations are misleading. A clear example of such events and their impact on an industrialized country, such as the United States, is the series of a total number of 15 natural disasters that occurred within the period from 2010 to July 2015. These hazards ranged from floods, hurricanes, and tornadoes to blizzards, wildfires, and mudflows, accounting for a total of 890 lost lives.

Both natural and man-induced fatalities might cause a variable number of victims. Nevertheless, natural disasters usually affect a great number of people and in most cases DNA-based identification is mandatory. Some of the most relevant circumstances surrounding natural disasters are the time elapsed since the moment the hazard hits a vulnerable area and the moment of retrieval of corpses and remains. In addition, the identification process might face serious restrictions, such as the lack of surviving relatives, which

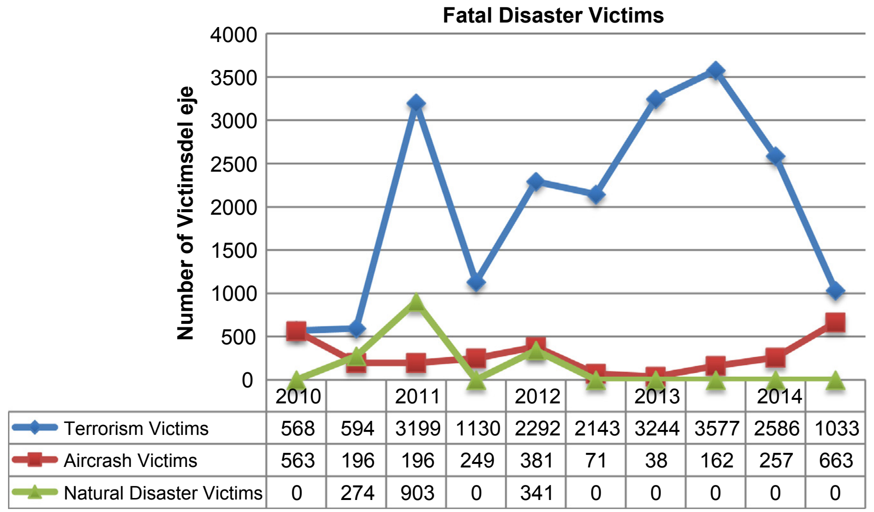
TABLE 22.1 Summary of Different Hazard Types and Their Effects on the Environment and Human Body

Promoting Force	Type of Hazard	Description	Type of Remains
Natural	Geological	Earthquake	Buried/decomposed
	Climatological	Volcanic eruption	Burnt
		Tsunami	Drowned/decomposed
		Hurricane	Drowned/decomposed
		Landslide	Buried/decomposed
	Biological	Epidemic	-
Man-induced	Technological	Environmental	-
		Engineering failure	Variable
		Transport accident	Fragmentary/decomposed
	Sociological	Terrorist attack	Fragmentary/decomposed
		War	Skeletonized

Catastropic Events

	2010		2011		2012		2013		2014	
Terrorist Attacks	40	20	239	67	131	112	98	81	174	47
Airplane Accidents	7	12	12	9	5	6	3	7	2	5
Natural Disasters	1	2	1	0	2	0	0	0	0	0

FIGURE 22.1 Worldwide evaluation of the total number of disaster events resulting from diverse disaster scenarios during 2010–2014.



	2010		2011		2012		2013		2014	
Terrorism Victims	568	594	3199	1130	2292	2143	3244	3577	2586	1033
Aircrash Victims	563	196	196	249	381	71	38	162	257	663
Natural Disaster Victims	0	274	903	0	341	0	0	0	0	0

FIGURE 22.2 Worldwide evaluation of the death toll resulting from diverse disaster scenarios during 2010–2014.

might otherwise have provided reference samples for comparison.

Moreover, man-induced fatalities might also widely vary in terms of the number of resulting victims. In some terrorist attacks, these victims may range from a few (tens/hundreds), as in the case of suicide bombings, to a considerable larger number (thousands), as in the event of the integrated terrorist attack that destroyed the World Trade Center in New York City on September 11, 2001.

The retrieval of remains is a time-consuming process, strongly influenced by the magnitude of the disaster and the degree of environmental modifications caused by the hazard. Accordingly, the fragmentary remains are usually inversely proportional to the dimension of the case, the number of victims, and the degree of environmental destruction. Hence, retrieval, cataloging, and storage represent the slowest steps in the analysis. Instead, DNA typing is considerably faster. DNA extractions can be carried out in a few hours with the help of automated platforms that may ensure a high throughput of molecular typing results. In most cases, living relatives can provide reference samples for comparison. Nevertheless, the victims' personal belongings are used in special cases where no relatives are available. Body fragmentation usually occurs in terrorist attack cases using explosives. Under these conditions, it is important to analyze every tissue fragment in order to reconstruct the body for further comparison with reference samples. In some cases, this strategy may provide

a clue concerning the number of victims resulting from the explosion and, potentially, to identify the suicide terrorist, if that is the case.

DNA-based identification is also mandatory in some special cases, such as those caused by transport accidents, in which the number and identity of the passengers are unknown but not all of them are deceased.

22.3 CONVENTIONAL IDENTIFICATION CRITERIA ROUTINELY USED FOR HUMAN IDENTIFICATION

The ability to recognize and identify human individuals represents a complex cognitive process. Although identification can be considered to be an automatic recognizing reflex, experienced by babies during their early months of life or even by nonhuman mammals such as pets, its underlying mechanisms are far from simple. The basic process requires an information storage device, a sensorial system that might allow the information to be captured and stored, and an adequate program that would allow for establishing coincidences between the previously stored information and the incoming data detected after the first capture. We practice this process daily unconsciously. This is true for a person's immediate recognition of his/her surrounding environment. This process is mostly subjective and is affected by memory loss or alterations in cognitive abilities.

Beyond the individual process, personal identification requires the use of objective criteria for developing robust and reliable identification systems that can contribute to personal identification at a nationwide environment. What kind of traits can provide useful information for an unbiased identification? Basically, these traits should be highly polymorphic and remain unchanged throughout life. Fingerprints are the only morphological attributes that may fulfill these requirements. Their use in identification can be traced back to the end of the 19th century when Francis Galton (1892) described the identification potential of finger ridges or dermatoglyphs.

During the 20th century, the creation of dactyloscopic databases, which are a part of identification documentation systems in most countries, was spread worldwide. Their identification power is remarkable since they can distinguish monozygous twins, a task that will fail if carried out by DNA analysis. Another criterion includes a person's signature, since calligraphic traces respond to an individual-specific pattern. Their analysis may correlate a piece of writing with a signature reference. Moreover, physiognomic traits are included as identification tools. However, it is evident that these features widely vary along a person's lifetime. Nevertheless, in a given life span they might provide useful identification information.

The abovementioned criteria were put together, and countrywide databases were created. These criteria represent the conventional identification approaches stored in "institutional memories" and depicted on our identification cards and passports, depending on the legal system of each country. These identification systems focus on establishing the identity of a given individual by comparing evidence, either a person's photographic image, a piece of writing, or even a latent fingerprint, with its corresponding reference, stored in a database. These criteria, which might be efficient for identifying an individual, are not suited for establishing biological relatedness.

In some special cases, such as mass disasters, aircraft crashes, terrorist bombings, mass floods, or earth slides, other information might be relevant in addition to dactyloscopic and photographic data available from databases. The following items, which can be provided by living relatives seeking missing family members, serve as an example: personal belongings, ante mortem information, including dental records, radiological studies revealing personal features, and a detailed description of personal marks, such as tattoos, piercings, or scars.

The systematic use of these kinds of identification parameters may be useful, but they require maximum care since the emotional pressure of relatives might lead to a biased recognition that could generate serious conflicts if the misidentification is detected at a later stage.

In addition to the so-called conventional criteria, mass disaster victim identification has been highly improved by the DNA-based identification approach. The first terrorist attack in which a wide set of DNA markers was used occurred in Argentina in 1992 (Corach et al., 1994), when the Embassy of Israel was hit by an explosive charge supposedly driven in a light truck by a suicide terrorist, whose remains were never found. The inclusion of this new identification approach called for the development of strategies for handling remains in order to ensure the optimization of DNA retrieval from samples, whenever possible.

22.4 CRITERIA FOR THE PRESERVATION OF REMAINS

The emergence of DNA-based identification methodologies in the forensic field brought about dramatic changes concerning sample preservation. Since their first use, restriction fragment length polymorphism (RFLP)- and PCR-based approaches have been applied to catastrophe-emerging remains. The first kind of markers mentioned above required the prevention of sample degradation, since polymorphic changes are detected within a wide range of DNA fragment sizes, strongly influenced by the single locus probes (SLPs) used and the restriction enzyme employed for DNA cleavage (mostly HaeIII with DNA

fragment sizes ranging from 1 to 10 kb). The second set of markers, STRs, were introduced by Edwards et al. (1991 and are widely spread in the forensic field. These markers can be informative even with partially degraded DNA samples; however, they are highly sensitive to interindividual contamination. Similar constraints are also valid for mitochondrial (mt) DNA sequencing of the D-Loop or Control Region that includes hypervariable regions I, II, and III (HVRI, II, and III) (Anderson et al., 1981; Vigilant et al., 1988; Andrews et al., 1999; Owen, 2000; Parson et al., 2004). Moreover, interspecies contamination has no effect on PCR-driven polymorphism detection, since species specificity is provided by the primer sequence used and is mostly human-specific, either for STRs or Control Region mtDNA amplifications.

In close relation to the described characteristics of the molecular approaches used for disclosing polymorphic attributes of the remains under study, practical procedures have been developed in order to optimize the material for preservation prior to analysis.

Mass fatality events occurring in Argentina were investigated by our group. During the Embassy case, the collection of remains was far from ideal, since there was insufficient prior knowledge available; therefore unintended errors were made. Inadequate handling of fragmentary body parts and long exposures to room temperature led to a certain degree of sample degradation and in some cases to interindividual contamination. These observations became apparent after DNA typing. The lessons we learned from this dramatic experience led us to develop guidelines on sample preservation procedures, whose implementation highly improved the results of other tragedies that were subsequently investigated.

These preservation procedures included the recovery of fragmentary remains in individual plastic bags and abundant tap and sterile water washing and rinsing prior to tissue sample collection. The tissues, mostly muscles, were placed in 50-mL sterile plastic tubes previously labeled and kept at −20°C until the moment of DNA extraction. These simple and basic rules allowed for the improved recovery of DNA samples, generally highly fragmented human remains obtained from the victims.

The second mass fatality event occurred 2 years later in Buenos Aires City, with more severe effects than in the previous case and producing a higher number of victims (Corach et al., 1995, 1996a). A terrorist attack using explosives completely destroyed a building and buried some of the victims. Ten months after the attack, when the building debris was removed completely, dehydrated muscle tissues were recovered. DNA was efficiently extracted, and its quality was suitable for STR typing or mtDNA sequencing. Based on the results obtained from the dehydrated soft tissue samples, a set of experiments was performed in order to develop a procedure that might warrant soft tissue preservation at room temperature. Since then, it has become our procedure of choice for fragmentary tissue preservation (Corach et al., 1996b). A few years later, when DNA quantitation by real-time PCR using a commercial quantification kit and genetic analysis "expert software" became available, a detailed analysis evaluating the preservation efficiency of dry table salt (commercial edible sodium chloride) on muscle tissue for 1 year at room temperature was carried out by our group. The sample preservation strategy based on salt dehydration proved to be highly efficient and allowed us to obtain complete genotypes with over 15 polymorphic STR loci and an amelogenin gene fragment (Caputo et al., 2011). At the moment of sample collection and for molecular identification purposes, the importance of complete sampling of all fragmentary remains must be stressed, even when some traits may lead to their visual identification. This practice will efficiently contribute to determining the number of victims and to reducing the number of unidentified remains at the morgue.

22.5 DNA POLYMORPHISMS USED FOR TRACING KINSHIP BETWEEN FRAGMENTARY HUMAN REMAINS AND THE RELATIVES CLAIMING THEM

During a short period between 1985 and 1992, a variety of diverse polymorphisms were identified, validated, and used in forensic casework investigation. RFLP analysis detected multilocus (Jeffreys et al., 1985) and single locus (Nakamura et al., 1987) minisatellites. PCR-based approaches included HLA DQ-A1 (Erlich and Bugawan, 1989) and Polymarker (Herrin et al., 1994) allele-specific oligotyping detected by reverse-dot blot hybridization; D1S80 PCR-amplified minisatellite (Budowle et al.,1991); STR typing (Edwards et al., 1991); and mtDNA sequence analysis (Ginther et al., 1992). Additionally, the inclusion of the markers harbored in the Y chromosome efficiently contributed to gender determination, which represented a relevant contribution to the identification of fragmentary remains.

With the exception of RFLP-based minisatellite DNA typing and the DQ-A1/Polymarker systems, all other molecular approaches, such as the automation of analytical platforms, either for sequence or fragment analysis, are currently in use with relevant improvements. Additional markers were included for investigating the fatalities of one of the most sadly conspicuous terrorist attacks: the crash of two planes into the North and South Towers of the World Trade Center, causing them to collapse on September 11, 2001, in New York City. The inclusion of single nucleotide polymorphisms (SNP) and the use of "mini" STRs highly improved the ability to

retrieve informative genetic profiles and genotypes from severely degraded samples.

22.6 CHALLENGES CONCERNING DNA DEGRADATION AND CONTAMINATION

Between the occurrence of a given disaster and the collection of remains, a highly variable period of time might elapse. Accordingly, the decomposition of corpses and human body fragments characterize the materials to be analyzed. In addition, the decaying process is strongly influenced by environmental conditions, such as temperature, humidity, soil, surrounding debris, etc. Decomposing is associated with macromolecular degradation, mostly hydrolysis and the deamination of proteins and the reduction of the molecular size of DNA fragments. The initially developed genetic polymorphic markers for human identification, although highly informative, proved to be of limited use when degraded samples needed to be analyzed. DNA integrity is also required for the evaluation of these markers, whose average fragment size should be as high as 10 kb. Furthermore, considerable amounts of genetic material are needed, due to the requirements imposed by the Southern blot hybridization approach. The efficient detection of hybridization signals needs at least 50 ng of DNA. Another limitation, determined by the nature of multilocus probes (MLPs) (Jeffreys et al., 1985) that are able to uncover highly polymorphic DNA minisatellites, is related to its specificity. These genome-wide distributed polymorphisms are present in almost every organism, including animals, plants, and fungi. Hence the lack of human specificity may obscure the hybridization patterns, since heterologous contaminant DNA may produce overlapping multiband results. MLP use was discontinued a few years after its discovery. However, its use as a probe for screening human genome libraries enabled the detection and isolation of a wide variety of human-specific clones, many of which denoted variable degrees of polymorphisms (Nakamura et al., 1987). These SLP markers were used as efficient probes that were sensitive to DNA integrity and quantity but unaffected by heterologous contamination. Their use in forensic identification was widespread, and some cases of mass fatalities were investigated with the help of these markers (Ludes et al., 1994; Corach et al., 1994). They were also used to confirm STR results obtained after the analysis of fragmentary remains.

Perhaps some of the most informative genetic markers ever developed for human identification were derived from SLP markers, in particular the MS31s, on which the repeat units denote detectable variation. These markers were termed minisatellite variant repeats (MVR) (Jeffreys et al., 1990). The combination of long-length PCR using primer-specific reactions followed by Southern blotting and hybridization with a specific SLP obtained highly informative genetic "bar codes." Their use in highly degraded samples was extremely efficient, combining SLP usefulness and PCR sensitivity. Their use allowed for the efficient identification of remains from the terrorist attack on a Jewish Mutual Association in Buenos Aires (Corach et al., 1996a). Unfortunately, these markers were not included in the routine casework marker battery.

Another efficient approach successfully used for obtaining valuable identification information from decomposed or skeletonized remains was provided by mtDNA sequence analysis. The most relevant attributes of this genetic marker reside in its size, structure, and multiplicity of copies present per cell. The small size of the mt genome (16,569-bp long) and its circular structure make it resistant to physical rupture and exonuclease degradation and the high copy number (with an average of five copies per organelle and over 1000 organelles/cells potential) allows retrieval of DNA fragments present in highly degraded samples. Maternal transmission of mtDNA only permits the matrilineage tracing of a given individual. The Control Region or D-Loop harbors a variety of polymorphic base substitutions and insertion/deletion polymorphisms within this noncoding region of about 1.2 kb. Three regions are recognized within it, denoted as HVRI (16,024−16,365), HVRII (73−349), and HVRIII (438−574). In some cases in which highly degraded material is analyzed, "mini-sequencing" strategies can be used in order to improve the results (Gabriel et al., 2001). This strategy is based on the amplification of small amplicons included in the hypervariable regions for further sequencing. The amplification of overlapping fragments allows for the "reconstruction" of badly degraded DNA fragments, since smaller fragments are more abundant in degraded samples. This approach has been successfully used in analyzing skeletonized human remains after long burial time periods (Corach et al., 1997).

As previously mentioned, forensic identification capabilities were boosted by the emergence of STR technology, which provided a powerful tool for routine use in establishing matches between suspects and crime scene evidences and investigating biological kinship relationships. Their main feature was to allow the typing of degraded samples containing minute DNA amounts. Moreover, the genome-wide distribution of STRs offers a chance to diversify their application field. The first set of STRs that appeared to be effective for forensic application was located on autosomal and X chromosomes (Edwards et al., 1991), and soon after, those located on the Y chromosome were described (Roewer et al., 1992). Autosomal STRs are fundamental for forensic evidence and individual identification; they are highly informative, can undergo recombination, and are inherited from both parents. Those located on the sex chromosomes have clearly different attributes. Those harbored on the X chromosome behave as autosomal markers in females with normal recombination, whereas in

males they fail to recombine with their counterpart, the Y chromosome. Accordingly, the Y chromosome is transmitted unrecombined from fathers to sons as a block of genetic information, and the sole variation they can incorporate is the product of mutation events. This character restricts their identification potential to the patrilineage level but provides a sensitive tool for detecting traces of male-specific material, such as semen in rape cases, and in the analysis of fragmentary remains. On the other hand, X chromosome STRs (X-STRs) have proved to be a relevant application in father-daughter cases (Schmidtke et al., 2004). The combined use of Y- and X-STRs has efficiently sped up the identification tasks performed in mass disaster case identification (Corach et al., 1995,1997).

Another additional improvement was introduced with the development of the "mini-STR." In these systems, the same loci are used. However, new primer sites were defined much closer to the variable region in which repeat units are tandemly organized. Accordingly, the amplification products are much more reduced in size, at 100 bp shorter than their regular STR counterparts (Wiegand and Kleiber, 2001).

SNP profiling is one of the most recent inclusions within the identification polymorphic systems. Since the number of alleles per system is restricted to two (biallele markers), the use of a complete set of SNPs may provide acceptable discrimination and exclusion powers (Gill, 2001). Since polymorphism is restricted to a single nucleotide, it offers the possibility of analyzing highly degraded DNA samples. Its usefulness was demonstrated in the September 11, 2001, terrorist attacks (Vastag, 2002).

22.7 CRITERIA EVOLUTION AND TECHNICAL APPROACHES APPLIED TO DNA-BASED VICTIM IDENTIFICATION IN MASS DISASTERS FROM THE EARLY 1990s TO DATE

As in many other fields of science, the continuous improvements in analytical processes, programming techniques, and software developments have allowed scientists to speed up data acquisition and processing. In addition, software design has contributed to optimizing the data-gathering systems included in automated sequencers and allowed for dealing with extremely high numbers of genotypes for rapid comparison/matching purposes. This led to the establishment of the National Database in Great Britain in 1995 and the Combined DNA Index System (CODIS) in the United States in 1997. In addition to their relevant use in forensic casework, the latter proved to be extraordinary helpful in the identification of victims emerging from the September 11, 2001, terrorist attacks (DNA President's initiative, 2004; http://www.dna.gov). However, many previous cases were investigated, and positive identification resulted from these analyses. The successful use of manual platforms was attained, despite being more laborious and time-consuming. Those labor-intensive approaches provided highly reproducible and robust results, but their efficiency for handling massive sample sets was limited.

In the early 1990s, the Southern blot hybridization approach allowed SLP typing, but few polymorphisms that could be detected by PCR-based methods were available. Most of these were at an experimental stage; however, in some special cases their use was widely justified. Additionally, genotype comparison, originally carried out by simple counting (Corach et al., 1995), was later replaced by simple programs that sped up the analytical process. This new approach made it possible to find samples that depicted genotype identities leading to body reconstruction or to establish possible kinship relationships between reconstructed bodies and putative relatives by comparing genotypes of remains/corpses and reference samples. Experimental strategies clearly varied due to technical feasibilities available at the time of each investigated case.

22.8 DESCRIPTION OF ANALYZED CASES

Different mass disaster situations are exemplified by six cases in which our lab was involved. These six mass disaster cases can be classified as man-induced disasters, and their occurrence can be defined by the following: (1) one closed accident, a commercial airplane explosion at approximately 2000 meters; (2) two open road accidents, one involving a commercial airplane crash immediately after takeoff and the other two buses that crashed and caught fire; and (3) three terrorist attack cases, two of them probably driven by religious motivations and one mass murder case conducted by state terrorism during the last military dictatorship in Argentina. A detailed description of these cases is discussed in the next paragraphs.

22.8.1 Terrorist Attack Cases

22.8.1.1 Terrorist Attack on the Embassy of Israel in Buenos Aires, Argentina

On March 17, 1992, a terrorist bomb completely destroyed the Israeli Embassy in Buenos Aires. Approximately 300 people were wounded, and the number of fatalities was estimated at 28, among them four Israeli members of the Embassy staff, four local embassy employees, some elderly residents at a nursing home, and school children on a passing bus. This attack marked the introduction of the Middle East (http://www.jewishvirtuallibrary.org/jsource/Terrorism/terrortoc.html) into South America and was the first mass disaster in which the remains were investigated using DNA-based technology in its early development stages.

Ten human remains were analyzed sometime after the explosion. The material was partially putrid, and intense bacterial contamination was denoted by the presence of 16 and 23S bacterial rRNA overimposed on some sample lanes. The analyses were carried out using single locus minisatellite YNH-24 (Nakamura et al., 1987) and six STRs, two of them located on the X chromosome (HUMARA and HUMHPRTB) and four on autosomal chromosomes (HUMTHO-1, FABP, RENA-4, and CD-4). Gender on the fragmentary remains was determined by amplifying the Yq13-ter sequences included within the heterochromatin region of the Y chromosome, described by Kogan et al. (1987). MtDNA sequencing of HVRI and HVRII was carried out by asymmetrical amplification.

Most samples were completely typed, and the data was compared with putative relatives; no missing victims were officially recognized. This was the first DNA analysis requested in Argentina by the Supreme Court of Justice in a criminal case.

22.8.1.2 Bombing of the Argentine Israeli Mutual Association in 1994

On July 18, 1994, the building of the Argentine Israeli Mutual Association was the bombing target of a terrorist action in the city of Buenos Aires. In this attack, 85 people were killed and hundreds were injured. It was Argentina's deadliest terrorist bombing aggression. The overall investigation required the analysis of over 350 samples, including human remains and reference samples. A rapid molecular screening was performed by employing five autosomal STRs (HUMTHO-1, FES/FPS, FABP, vWA, HUMRENA4) and two sex chromosome STRs: HUMHPRTB on the X chromosome and Y27H39 on the Y chromosome, also known as DYS19. STR profiling results were confirmed by SLP typing in those cases in which molecular integrity allowed the retrieval of quality-suited DNA for the analysis. Four minisatellite probes (YNH-24, PH-30, MS1, and LH-1) were used in subsequent hybridization/dehybridization cycles employing a single Southern blotted membrane for each set of samples. In those cases with highly degraded DNA, confirmation was attained by the PCR-MVR approach (Jeffreys et al., 1990) or mtDNA HVRI and HVRII sequencing.

22.8.1.3 Mass Murder Case Conducted by State Terrorism During the Last Military Dictatorship in Argentina

Argentina was ruled by military forces during the period between 1976 and 1983. During this time, the bloodiest repression ever was perpetrated by state terrorism. Over 30,000 people were supposedly killed or missing, of which over 11,000 disappearances have been well documented. Diverse kidnapping/torture/killing modalities were used, which resulted in missing victims whose corpses were never retrieved. However, anonymous skeletal remains were discovered in public cemeteries, known as "NN burial sites." In order to investigate the efficiency of the molecular typing approaches, sets of bones and teeth were analyzed with several typing approaches, including the nested-PCR reaction using external primer sequences to some STRs, mainly HUMTHO-1 and HUMFABP, and mtDNA sequencing. Although the first approach produced reproducible results in some samples, mtDNA sequencing proved to be successful in all samples tested. Unfortunately, at that time no sequence or genetic profile databases were available for comparison, which would have led to missing person identification (Corach et al., 1997).

22.8.2 Open Road Accidents

22.8.2.1 Bus Crash Accident

On January 9, 1993, two buses crashed, and one of them caught fire. The accident caused 56 fatal victims and a total of 80 injured passengers. The accident occurred on national route 14 near Santo Tome in the Province of Corrientes, Argentina. Two years later several people were still missing, who had probably been victims of the multiple bus crash. A DNA investigation was requested, for which a combination of genetic markers, including autosomal STRs, Y-STRs, and mtDNA HVRI and HVRII were characterized. A set of 25 human fragmentary remains, including badly burned bones and teeth, were analyzed, along with the reference samples obtained from relatives. Considering the quality of the analyzed remains, DNA extraction and typing proved to be successful in 65% of the cases (16 out of 25), including the missing son of one couple. In this latter case, complete genotypes were obtained with four autosomes, one X-STR, and nine Y-STRs. The analytical strategy was focused on Y-STR haplotype detection of the putative son with the haplotypes obtained from the fragmentary remains. A premolar revealed the same haplotype and also shared at least one autosomal allele with both descendants and with the putative mother. HVRI and HVRII confirmed the matrilineage origin since the putative mother presented an identical mt haplotype.

22.8.2.2 Open Road Airplane Crash

On August 31, 1999, LAPA Flight 3142 from Buenos Aires to the capital city of Cordoba, Argentina, crashed near Jorge Newbery airport, shortly after takeoff. The crash resulted in 65 fatalities, 40 survivors, and 17 seriously injured passengers, which makes it one of the deadliest accidents in the history of Argentine aviation. Due to human errors and technical failure, the aircraft retracted flaps

during takeoff, and even though the aircraft had achieved minimum takeoff speed, it continued beyond the ramp, failing to get airborne, broke through the airport's fence, crossed a road, dragged with it an automobile, and finally collided against road construction machinery. Gasoline spilling over the hot engines caused ignition and the total destruction of the aircraft.

A total of 38 human remains were received for DNA typing, and the members of 13 family groups provided reference samples. DNA was extracted from corpse material in triplicate and also blindly analyzed in triplicate. The analytical strategy focused on autosomal STR typing and gender determination by using amelogenin. All detection analyses were carried out with automated sequencers, and 13 autosomal STRs were typed with commercial kits, while Y-STRs were analyzed with custom-made multiplexes using the minimal haplotype marker system (Kayser et al., 1997). After identifying male samples by amelogenin gender-associated dimorphism, patrilineage tracking was conducted by Y-STR haplotyping of suitable reference samples.

Nine remains were identified as belonging to individuals that were biological relatives of reference samples' donors. Four of them did not correspond to any family group. Unfortunately, bodies were prematurely returned to relatives, and hence errors occurred, forcing a complete retyping of exhumed bodies. This dramatic situation underscored the remarkable value of DNA typing not only for identifying remains and individuals but also for ensuring the proper use of other conventionally employed forensic identification practices, such as visual recognition, personal belongings, odontology, and radiology, among others.

22.8.3 Closed Accidents

22.8.3.1 In-Flight Airplane Explosion

On May 18, 2011, passenger flight 5428 of the Argentine operating company Sol Airlines exploded at an altitude of about 2000 m in Prahuaniyeu, Río Negro Province, Argentina. The plane, a Saab 340A, was operating a domestic flight from Presidente Perón International Airport, Neuquén, to General Enrique Mosconi International Airport, Comodoro Rivadavia, Chubut Province. The aircraft came down 25 km southwest of the town of Los Menucos, Río Negro. A report suggested that it was caused by severe icing of the airframe. The black box was retrieved 2 days later near the disaster area. All 22 people aboard, 3 crew members and 19 passengers, were killed. Victims were highly fragmented, and 418 specimens were recovered for DNA typing.

Efficient corpse tissue preservation, automated DNA extraction, sensitive DNA quantification, and expert analysis software make it possible to design highly efficient workflows that speed up the mass disaster victim identification process from highly fragmented human bodies.

The victims' identifications were based on genetic profile comparisons between fragmentary remains and first-degree relatives; no personal belongings were needed since every victim was claimed by potential biological relatives. Genetic profiles combined genotypes and haplotypes obtained by autosomal and Y-STR analysis, respectively. All bodies were severely fragmented, and a large number of isolated cranial bones were found. Remains were collected at the disaster area and sent to Buenos Aires where samples were selected and preserved in 50-mL polypropylene tubes containing solid sodium chloride (table salt). DNA extractions from muscle and bone were performed using a semiautomated DNA purification system; quantification was carried out by real-time PCR, STR amplification using commercial kits, and STR profile analysis with the help of expert software. The implementation of this strategy allowed us to identify all the victims in one week. However, the complete task of identifying all remains recovered within the disaster area took over 10 months due to the wide dispersion of fragments and the court's decision to identify each morphologically recognizable human tissue found within the disaster area (Corach et al., 2013).

22.9 FROM FORENSIC GENETICS TO FORENSIC GENOMICS: THE CHANGE OF A PARADIGM DRIVEN BY TECHNOLOGY

For over 20 years, autosomal fluorochrome-label STRs were the gold standard for human identification purposes, representing the most powerful tool for forensic identification.

In addition to the subtle improvements introduced within the analytical platforms, automated sequencers, the most important modification introduced to the identification toolbox, involved the number of polymorphic STRs employed and the amplification efficiency due to improved amplification cocktails, denoting higher sensitivity and more resistance to the effects of inhibitors.

The 13 STRs included within CODIS launched in 1997 will be increased by the addition of seven markers. The new 20 STR FBI CODIS core set will be fully implemented on January 1, 2017. The increasing number of markers represents the mandatory need to avoid the detection of fortuitous matches resulting from the great number of profiles hosted within the National DNA Index System.

The apparent technological stability of canonically admitted markers contrasts with the new forensic tools developed as a result of genomic research. In a comprehensive review, Kayser and de Knijff (2011) describe a wide span of new analytical approaches that open

unexpected perspectives for forensic research. Many of the new developments will effectively contribute to the individual identification of fragmentary remains from disastrous situations that might occur in the future or that occurred in the past, such as mass graves or collective burial sites produced in the course of past violent conflicts. The new tools include the identification potential of highly degraded DNA (<100 bp-long templates) by SNP. Moreover, NGS platforms also offer the possibility of determining phenotypical traits from highly fragmented genetic material. The opportunity for gathering information concerning eye, hair, and skin color of the people whose remains are being investigated might highly contribute to the identification process, as shown in the identification of King Richard III of England (King et al., 2014). The new NGS platforms enable phenotypic and ancestral/geographic prediction and allow for the genotyping of the canonical STRs included in the intelligence databases developed worldwide, most of them under CODIS principles. The intrinsic usefulness of the STRs used since 1997 is now available by sequencing rather than by analyzing the fragment length of amplified alleles. Since the most informative STRs used are either complex or compound, if we restrict analysis to their fragment length, their informative content diminishes. STRs exhibit a differential structural organization, and those used in forensic analysis are included in four categories. Simple STRs show identical repeat units in length and sequence; their informative content is identical if analyzed by its length or by sequencing. Compound STRs include two or more adjacent simple repeat motifs, whereas complex STRs include several repeat blocks of variable unit length as well as variable intervening sequences. Finally, complex hypervariable STRs are challenging due to their complex molecular structure, although one such STR has been included in commercial kits (SE33). The NGS platforms designed for forensic identification combine canonical markers, increasing their informative content, and provide additional valuable information that may be relevant in disaster victim identification scenarios.

22.10 FUTURE PERSPECTIVES

Soon after the beginning of the new century, complex and multitudinous terrorist-produced mass disasters occurred. These challenging situations have forced forensic scientists to optimize and update the identification approaches based on DNA typing. The most dramatic terrorist attacks have taken place in considerably recent times. The most relevant ones are the September 11 suicide attacks on the Twin Towers at the World Trade Center in New York City in 2001, which killed 2974 people; the terrorist attack perpetrated against Atocha and other Madrid City railway stations on March 11, 2004; and the bombing attack on London's public transport system on July 7, 2005, which killed 55 people. In addition to these man-induced disasters, there are those that were naturally produced, such as the tsunami in Southeast Asia in December, 2004; Hurricane Katrina in the United States in August, 2005; Cyclone Nargis that hit Myanmar on May 3, 2008, an extreme catastrophe causing over 80,000 fatalities; and the destructive earthquake in China on May 15, 2008, whose final death toll is 69,136 victims.

In concordance with the magnitude of these catastrophes that might require DNA-based identification approaches, multilaboratory facilities are needed. Additionally, international organizations such as Interpol have established detailed protocols for mass disaster victim identification, and the International Commission on Missing Persons may also provide advisory and technical help for mass fatalities worldwide. High-throughput platforms are efficiently powering a modern investigation of such disasters. Molecular diagnosis based on DNA-typing approaches has undergone an intense evolutionary process that might efficiently respond to the intensely demanding victim identification needs imposed by extreme mass catastrophes that unfortunately characterize this age.

REFERENCES

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J., Staden, R., Young, I.G., 1981. Sequence and organization of the human mitochondrial genome. Nature 290, 457−465.

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., Howell, N., 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat. Genet. 23, 147.

Budowle, B., Chakraborty, R., Giusti, A.M., Eisenberg, A.J., Allen, R.C., 1991. Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. Am. J. Hum. Genet. 48 (1), 137−144.

Corach, D., Penacino, G., Sotelo, A., 1994. Dealing with human remains in disaster areas. The case of the Israeli Embassy explosion occurred in Buenos Aires. Adv. Forensic Haemogenet. 5, 259−261.

Caputo, M., Bosio, L.A., Corach, D., 2011. Long-term room temperature preservation of corpse soft tissue: an approach for tissue sample storage. Investig. Genet. 2011 (2), 17.

Corach, D., Sala, A., Penacino, G., Sotelo, A., 1995. Mass disasters: rapid molecular screening of human remains by means of STRs typing. Electrophoresis 16, 1617−1623.

Corach, D., Sala, A., Penacino, G., Sotelo, A., 1996a. Combined STR, VNTR and MVR typing and mtDNA sequencing led to identification of human remains emergent from the AMIA explosion in Buenos Aires. Adv. Forensic Haemogenet. 6, 337−339.

Corach, D., Penacino, G., Sala, A., 1996b. Room temperature preservation of putrid cadaveric soft tissues by means of inorganic compounds allows efficient recovery of DNA. In: Proceedings 7th International Symposium on Human Identification, p. 203.

Corach, D., Penacino, G., Sala, A., Iannucci, N., Bernardi, P., Doretti, M., Fondebrider, L., Ginarte, A., Inchaurregui, A., Somigliana, C., Turner, S., Hagelberg, E., 1997. Additional approaches to DNA typing of skeletal remains: the search for missing persons killed during the last dictatorship in Argentina. Electrophoresis 18, 1608−1612.

Corach, D., Sala, A., Bobillo, C., Caputo, M., Alechine, E., Irisarri, M., Marino, M., Canonaco, E., Rodriguez, C., 2013. Optimized mass fatalities victim identification: airplane crash as a test case. Forensic Sci. Int. Genet. Suppl. Ser. 4, e222−e223.

Edwards, A., Civitello, A., Hammond, H.A., Caskey, C.T., 1991. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am. J. Hum. Genet. 49, 746−756.

Edwards, A., Hammond, H.A., Jin, L., Caskey, C.T., Chakraborty, R., 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12, 241−253.

Epplen, J.T., McCarrey, J.R., Sutou, S., Ohno, S., 1982. Base sequence of a cloned snake W-chromosome DNA fragment and identification of a male-specific putative mRNA in the mouse. Proc. Natl. Acad. Sci. U. S. A. 79 (12), 3798−3802.

Erlich, H.A., Bugawan, T., 1989. HLA Class II gene polymorphism: DNA typing, evolution, and relationship to disease susceptibility. In: PCR Technology. Principles and Applications for DNA Amplification. Stockton, New York, pp. 193−208.

Galton, F., 1892. Hereditary Genius: An Inquiry Into Its Laws and Consequences. Macmillan, Fontana, London.

Gabriel, M.N., Huffine, E.F., Ryan, J.H., Holland, M.M., Parsons, T.J., 2001. Improved MtDNA sequence analysis of forensic remains using a "mini-primer set" amplification strategy. J. Forensic Sci. 46, 247−253.

Gill, P., 2001. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. Int. J. Leg. Med. 114, 204−210.

Ginther, C., Issel-Tarver, L., King, M.C., 1992. Identifying individuals by sequencing mitochondrial DNA from teeth. Nat. Genet. 2, 135−138.

Herrin, G., Fildes, N., Reynolds, R., 1994. Evaluation of the AmpliType® PM DNA test system on forensic case samples. J. Forensic Sci. 39, 1247−1253.

Jeffreys, A.J., Wilson, V., Thein, S.L., 1985. Hypervariable "minisattelite" regions in human DNA. Nature 314, 37−73.

Jeffreys, A.J., Neumann, R., Wilson, V., 1990. Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. Cell 60, 473−485.

Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., Penacino, G., Perez-Lezaun, A., Piccinini, A., Prinz, M., Schmitt, C., Schneider, P.M., Szibor, R., Teifel-Greding, J., Weichhold, G., de Knijff, P., Roewer, L., 1997. Evaluation of Y-chromosomal STRs: a multicenter study. Int. J. Leg. Med. 110, 124−125.

Kayser, M., de Knijff, P., 2011. Improving human forensics through advances in genetics, genomics and molecular biology. Nat. Rev. Genet. 12, 179−192.

King, T., Gonzalez Fortes, G., Balaresque, P., Thomas, M., Balding, D., Maisano Delser, P., Neumann, R., Parson, W., Knapp, M., Walsh, S., Tonasso, L., Holt, J., Kayser, M., Appleby, J., Forster, P., Ekserdjian, P., Hofreiter, M., Schürer, K., 2014. Identification of the remains of King Richard III. Nat. Commun. 5, 5631. http://dx.doi.org/10.1038/ncomms6631.

Kogan, S.C., Doherty, M., Gitschier, J., 1987. An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. Application to hemophilia A. N. Engl. J. Med. 317, 985−990.

Landstainer, K., 1901. Ueber Agglutinationserscheinungen normalen menschlichen Blutes. (Translation: on Agglutination phenomena of normal human blood). In: Boyer, S.H. (Ed.), 1963, Papers on Human Genetics, pp. 27−31. Wien. Klin. Wochenschr. 14, 1132−1134.

Ludes, B., Tracqui, A., Pfitzinger, H., Kintz, P., Levy, F., Disteldorf, M., Hutt, J.M., Kaess, B., Haag, R., Memheld, B., Kaempf, C., Friederich, F., Evento, E., Mangin, P., 1994. Medico-legal investigations of the Airbus, A320 crash upon mount Ste-Odile, France. J. Forensic Sci. 5, 1147−1152.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., Erlich, H., 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb. Symp. Quant. Biol. 51, 263−273.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E., White, R., 1987. Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 235, 1616−1622.

Owen, R., 2000. Karl Landsteiner and the first human marker locus. Genetics 155, 995−998.

Panicker, S.G., Singh, L., 1994. Banded krait minor satellite (Bkm) contains sex and species-specific repetitive DNA. Chromosoma 103 (1), 40−45.

Parson, W., Brandstätter, A., Alonso, A., Brandt, N., Brinkmann, B., Carracedo, A., Corach, D., Froment, O., Furac, I., Grzybowski, T., Hedberg, K., Keyser-Tracqui, C., Kupiec, T., Lutz-Bonengel, S., Mevag, B., Ploski, R., Schmitter, H., Schneider, P., Syndercombe-Court, D., Sørensen, E., Thew, H., Tully, G., Scheithauer, R., January 28, 2004. The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: organisation, results and perspectives. Forensic Sci. Int. 139 (2−3), 215−226.

Roewer, L., Amemann, J., Spurr, N.K., Grzeschik, K., Epplen, J.T., 1992. Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. Hum. Genet. 89, 389−394.

Schmidtke, J., Kühnau, W., Wand, D., Edelmann, J., Szibor, R., Krawczak, M., 2004. Prenatal exclusion without involving the putative fathers of an incestuous father-daughter parenthood. Prenat. Diagn. 24, 662−664.

Tautz, D., Renz, M., 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res. 12 (10), 4127−4138.

Vastag, B., 2002. Out of tragedy, identification innovation. JAMA 288, 1221−1223.

Vigilant, L., Stoneking, M., Wilson, A.C., 1988. Conformational mutation in human mtDNA detected by direct sequencing of enzymatically amplified DNA. Nucleic Acids Res. 16, 5945−5955.

Watson, J.D., Crick, F.H., 1953. A structure for deoxyribose nucleic acid. Nature 171, 737−738.

Weber, J.L., Mayr, P.E., 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. 44, 388−396.

Wiegand, P., Kleiber, M., 2001. Less is more length reduction of STR amplicons using redesigned primers. Int. J. Leg. Med. 114, 285−287.

Wyman, A.R., White, R., 1980. A highly polymorphic locus in human DNA. Proc. Natl. Acad. Sci. U.S.A. 77, 6754−6758.

Wisner, B., 2004. Assessment of capability and vulnerability. In: Bankoff, G., Frerks, G., Hilhorst, D. (Eds.), Mapping Vulnerability Disasters, Development and People. Earthscan, London, pp. 183−193.

Chapter 23

Preimplantation Genetic Diagnosis

M. De Rycke and C. Staessen
Universitair Ziekenhuis & Vrije Universiteit Brussel (VUB), Brussels, Belgium

23.1 INTRODUCTION

Preimplantation genetic diagnosis (PGD) is performed for couples at a high risk of transmitting a known genetic condition to their offspring. It is a well-established alternative to prenatal diagnosis, involving the biopsy and genetic testing of a single or multiple cells from in vitro-obtained oocytes and/or preimplantation embryos. As only embryos shown to be unaffected for the genetic condition under study are selected for transfer to the patient's uterus, PGD offers the advantages of circumventing an invasive prenatal diagnosis and therapeutic abortion. PGD requires an in vitro fertilization (IVF) treatment; this treatment with additional risks and costs may be considered to be a burden for fertile couples. Couples at risk of transmitting a genetic condition and concurrently suffering from infertility may more easily combine IVF and PGD.

The technology applied in preimplantation genetic screening (PGS) or aneuploidy screening is similar to PGD, but the objective is different. PGS involves the selection of euploid embryos to improve IVF results and to reduce the chance of pregnancies with chromosomal abnormalities (Verlinsky and Kuliev, 1996). It has been well established that human embryos carry chromosomal abnormalities in high proportions. Taking this fact together with an increased aneuploidy risk at advanced maternal age and the findings of high aneuploidy rates in spontaneous abortions, PGS has been offered to specific IVF patient groups with an increased risk for embryos with chromosomal abnormalities, most commonly patients of advanced maternal age and patients with recurrent IVF failure or repeated miscarriages (not due to translocations).

PGS is still widely applied, despite the controversy about whether it is clinically useful or not. Several randomized controlled trials failed to show a benefit for PGS using fluorescent in situ hybridization (FISH) for 5 to 12 chromosomes and cleavage-stage biopsy (Checa et al., 2009). The European Society of Human Reproduction and Embryology (ESHRE) PGD consortium issued a position statement about this in 2010 and recommended that alternative approaches should be evaluated, that is, alternative stages at which to biopsy and carry out genetic analysis as well as alternative, genome-wide methods for the analysis (Harper et al., 2010). PGD centers have adapted their methods since then, but the routine clinical implementation still awaits reliable data from prospective clinical trials. As PGS is offered to (sub)fertile couples not carrying a genetic disease, it will not be further discussed in this chapter.

PGD developed in the wake of human IVF and polymerase chain reaction (PCR) technology. The first report on children born after the invention of PGD was by Handyside et al. in 1990, describing the use of PCR for the detection of repetitive Y-sequences for gender determination in families with X-linked diseases. Later, FISH was adapted to single cell interphase nuclei, and it became the standard method for sexing in X-linked diseases as well as for inherited chromosomal aberrations. Single cell PCR has been the method of choice for the detection of single gene defects or monogenic disorders (dominant or recessive, autosomal, or X-linked). The gold standard methods of FISH and PCR are being replaced by whole-genome amplification (WGA) and genome-wide technologies.

PGD involves a multidisciplinary team of professionals in IVF, genetics, ethics, and psychology. A close collaboration is required between the IVF unit and a genetic diagnosis unit, specialized in genetic analysis for PGD. Both units may be physically and organizationally within the same institute. Alternatively, transport PGD is carried out, which involves the biopsy of embryo samples in a satellite IVF unit followed by transportation, often over a long distance, to the genetic diagnosis unit.

Guidelines for best practices have been drawn up by the Preimplantation Genetic Diagnosis International Society

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00023-7
Copyright © 2017 Elsevier Ltd. All rights reserved.

(Preimplantation Genetic Diagnosis International Society (PGDIS), 2008). In addition, four extensive guidelines on different aspects of PGD (organization of a PGD center, FISH-based testing, amplification-based testing, and biopsy) were established by the ESHRE PGD Consortium (Harton et al., 2011a–d).

23.2 ASSISTED REPRODUCTIVE TECHNOLOGY AND BIOPSY

23.2.1 In Vitro Fertilization

Assisted reproductive technology (ART) is an intrinsic part of PGD. With the exception of the procedure for oocyte/zygote or embryo biopsy, ART within PGD is essentially the same as for infertility treatment, irrespective of whether the couple is fertile or not. Controlled ovarian hyperstimulation supports the production of multiple mature oocytes to create sufficient zygotes/embryos and increase the chance of identifying embryos with the desired genetic status. All cumulus cells should be removed from the aspirated oocytes. Furthermore, fertilization should be through intracytoplasmic sperm injection rather than regular IVF. Remaining cumulus or residual sperm cells adhering to the zona pellucida (ZP) may lead to contamination when DNA amplification is used for diagnosis.

The biopsy procedure involving the perforation of the ZP followed by removal of one or several cells for genetic analysis can be carried out at different developmental stages (see Sections 23.2.2–23.2.4). The ZP opening is achieved mechanically, by cutting through the ZP with a microneedle; chemically, by dissolving part of the ZP with acidic Tyrode's solution; or, in the majority of cases, through modulating a laser beam via the optical system of a microscope. Cell removal is then carried out by extrusion or aspiration (De Vos and Van Steirteghem, 2001; Xu and Montag, 2012). After biopsy, cells are washed and either fixed on a slide for FISH analysis or collected in small tubes for subsequent amplification-based testing.

23.2.2 Polar Body Biopsy

Polar bodies (PBs) are produced in the first and second meiotic division as oocytes complete maturation upon fertilization. PB analysis represents an indirect method in which the genotype or chromosomal constitution of the oocyte is derived from the complement present in the PBs. An accurate diagnosis should be based on an analysis of both the first and second PBs to preclude a misdiagnosis, which may arise from recombination or allele dropout (ADO) (monogenic diseases) or by nondisjunction or meiotic errors (chromosomal analysis). The doubling of the total sample number together with the fact that some oocytes fail to fertilize or form embryos may be regarded as a waste of time and resources unless the PB samples can be stored, and only those corresponding with normal embryo development are ultimately processed.

For PB biopsy, the ZP is breached using either mechanical slitting with a fine needle or laser technology, as the acidic Tyrode's solution may adversely affect subsequent oocyte development. Although both PBs may be biopsied simultaneously (requiring only one manipulation), it may be difficult to differentiate them morphologically, and sequential biopsy has been recommended to ensure a correct PB identification (Treff et al., 2012).

The advantages of PB biopsy are self-evident: as PBs do not contribute to normal fertilization or embryonic development, their removal has no detrimental effects. In countries where genetic testing should be finished before syngamy, PB biopsy is the only legal option, but it leaves very little time to complete diagnosis. Conversely, if legal restrictions or ethics are not an issue, more time is available for analysis compared with a cleavage-stage biopsy. The most important limitation of PB biopsy is that only the maternal genetic contribution can be evaluated.

23.2.3 Cleavage-Stage Biopsy

Cleavage-stage biopsy has been used in the majority of PGD cycles during the last couple of decades. One or two blastomeres are biopsied on the morning of day 3, when the embryo is normally at the six- to eight-cell stage. It is assumed that all cells are still totipotent, and compaction has not yet occurred. In the event that compaction has started, a brief exposure of the embryos to Ca^{2+}/Mg^{2+}-free media will reduce the adherence between cells, facilitating blastomere removal. Mechanical, chemical means, and laser are all suitable for zona breaching when performing cleavage-stage biopsy (Harton et al., 2011d; Xu and Montag, 2012). PGD may be based on the analysis of either one or two biopsied blastomeres from a single embryo. The impact of the removal of one or two cells on further embryonic development and implantation potential is highly debated. De Vos and coworkers showed that a two-blastomere removal at the cleavage stage harms embryonic development more than the removal of only one blastomere. Within their prospective cohort of single embryo transfers they also demonstrated that the live birth rate after a one-cell removal from eight-cell embryos (37%) was similar to the rate of a control intracytoplasmic sperm injection (ICSI) group without biopsy (35%) but significantly higher than the rate of the two-cell biopsy study group (22%) (De Vos et al., 2009). Another paired clinical trial showed that implantation rates of 50% in the nonbiopsied group were diminished to 30% in the

cleavage-stage biopsied group, while for the nonbiopsied versus blastocyst-biopsied groups similar implantation rates were obtained (Scott et al., 2013). Therefore the removal of a single blastomere at the cleavage stage is recommended, provided that an accurate and reliable single cell test is available.

The major disadvantage of cleavage-stage biopsy is the limited amount of DNA present in a single blastomere for analysis. In addition, high rates of mosaicism are observed in embryos at this early stage of development, although for monogenic disorders, errors attributed to false positive or false negative results may be minimized by applying appropriate diagnostic strategies and interpretation of results. The clear advantages of cleavage-stage biopsy are that genetic diseases of both paternal and maternal origin can be detected and that it leaves sufficient time for analysis before transfer on day 5. The transfer procedure was originally carried out on day 3 or day 4, but the change to sequential embryo culture media has allowed for an extended embryo culture and the transfer of only those embryos reaching the blastocyst stage (Gardner and Lane, 1998). Fig. 23.1 shows an image of a cleavage-stage and a blastocyst biopsy.

FIGURE 23.1 Cleavage stage (A) and blastocyst (B) biopsy. The embryo is held in a fixed position with a holding pipette (left), and a smaller biopsy pipette is used to remove cells through an opening in the zona pellucida, which was created with laser pulses. One or two nucleated blastomeres are removed on day 3, whereas on day 5 a small group of cells is taken from the trophectoderm layer of the blastocyst.

23.2.4 Blastocyst Biopsy

The blastocyst is the highest degree of development that an embryo can reach in vitro. It develops about 5–6 days postinsemination and it is characterized by cells from the inner cell mass (ICM), the outer trophectoderm (TE) cell layer, and the blastocoel, which is a cavity filled with fluid containing high concentrations of lactate and specific amino acids and lower concentrations of glucose and pyruvate.

The feasibility of TE biopsy was first shown by Dokras et al. (1990), and the first successful clinical applications for PGD were reported in 2005 (McArthur et al., 2005). Zona breaching may be performed by mechanical means or by laser in the TE region opposite from the ICM, and herniating TE cells are cut from the embryo on day 5 or day 6. TE biopsies typically contain about five cells, which is an advantage as this will improve diagnostic accuracy. As the ICM from which the embryo proper will develop is kept intact, TE biopsy is considered to be a safer option than cleavage-stage biopsy. Data suggest that TE biopsy has no adverse impact on embryonic developmental competence (Scott et al., 2013), and comparison even suggests that TE biopsy may yield higher implantation and live birth rates than cleavage-stage biopsy. The biological phenomenon of chromosomal mosaicism is also present at this developmental stage, and it may distort overall concordance between biopsied samples and embryos per se, but several reports indicate that the mosaicism level at the blastocyst stage is lower than at the cleavage stage. The problem of a limited time for analysis in the case of fresh embryo transfer (ET) at day 5/6 can be overcome by cryopreservation (relying on efficient vitrification and thaw-survival protocols) and ET in a deferred cycle. The decision to opt for this latter strategy enables a more efficient and cost-effective laboratory organization, as the larger time windows make it possible to pool and coprocess samples of multiple patients. Moreover, there is at present sufficient evidence demonstrating that the deferred transfer to a "nonstimulated" cycle has its advantages in terms of implantation, ongoing pregnancy rates, and a lower risk of genetic and epigenetic alterations (Roque et al., 2013). Fig. 23.2 represents an overview of a PGD cycle with biopsy on day 3 (A) or biopsy on day 5/6 (B).

As an alternative for TE cells, studies investigated the use of DNA isolated from the blastocoel fluid (BF), which was aspirated from expanded blastocysts with a micropipette. So far, a few small and two larger studies with contradictory outcomes have been published; one study found a high karyotype concordance between BF and TE (Gianaroli et al., 2014), but a second study showing a

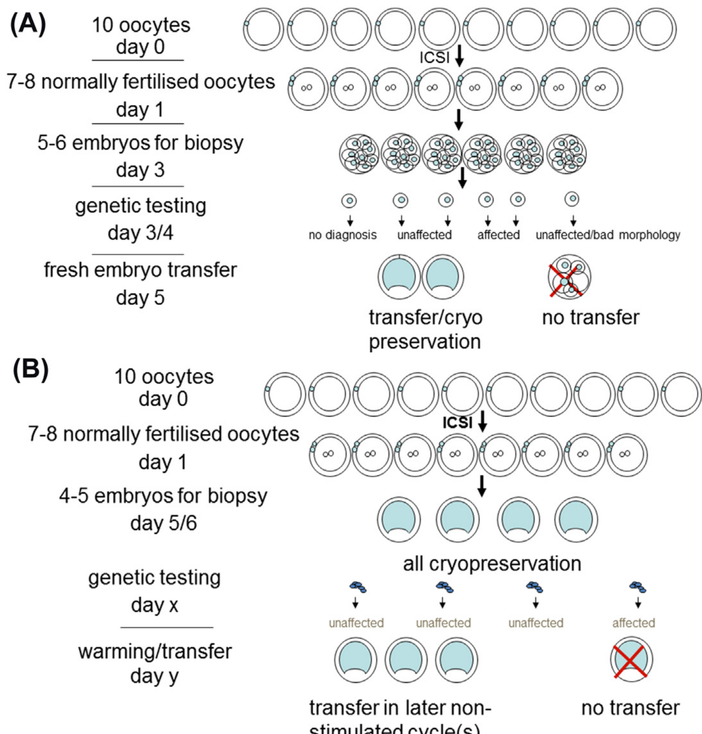


FIGURE 23.2 Scheme for a preimplantation genetic diagnosis cycle with a biopsy on day 3 (A) versus a preimplantation genetic diagnosis cycle with a biopsy on day 5/6 (B). In both schemes, oocyte collection and intracytoplasmic sperm injection (ICSI) are carried out on day 0. On average, 70–80% of the injected oocytes are normally fertilized. On day 3, 50–60% of the embryos are suitable for cleavage-stage biopsy, whereas on day 5/6 40–50% of the embryos are suitable for blastocyst biopsy. On day 3, one or two cells are biopsied and genetically tested; the results are available in time to allow for a fresh embryo transfer on day 5. Embryos without a diagnosis and affected embryos are not eligible for transfer. Genetically transferable embryos of a good morphological quality can be transferred; the number of embryos depends on several factors (female age, rank of trial). Supernumerary embryos are cryopreserved. On day 5/6, 5–10 cells are biopsied and genetically tested; the time window is too short, and all embryos are cryopreserved while the genetic testing is pending. The genetically transferable embryos are transferred in a later, nonstimulated cycle.

discordant karyotype between BF and ICM–TE in 52% of embryos questions whether BF biopsies are truly representing the embryo. The high aneuploidy level within BF from euploid blastocysts in the latter study suggests a correction mechanism with a clearance of aneuploid cells into the blastocoel that may underlie the gradual decreased ratio of aneuploid cells from the cleavage stage to blastocyst (Tobler et al., 2015). The concept of BF samples for genetic testing in PGD/PGS requires further exploration before clinical application.

23.3 PREIMPLANTATION GENETIC DIAGNOSIS FOR MONOGENIC DISORDERS

23.3.1 Indications

PGD was initially applied for the same indications as prenatal diagnosis. PGD has also been applied for indications for which prenatal diagnosis is ethically difficult, such as cancer predisposition syndromes and other late-onset diseases (Woodson et al., 2014; Derks-Smeets et al., 2014).

Not surprisingly, most PGD cycles have been reported for the more frequently occurring disorders: cystic fibrosis, spinal muscular atrophy and beta-hemoglobinopathies for autosomal recessive disorders, myotonic dystrophy type 1, neurofibromatosis, and Huntington's disease for the autosomal dominant disorders (Harper et al., 2012). For the X-linked disorders, PGD is mainly carried out for Duchenne's muscular dystrophy, hemophilia, and fragile X syndrome. The advantages of specific DNA diagnosis over sexing for X-linked disorders are 2-fold: healthy male embryos are not discarded, and female carrier embryos can be identified and possibly used for transfer, according to the patient's wishes and the center's policy.

A special indication is Human Leukocyte Antigen (HLA) typing of preimplantation embryos to select an embryo that is HLA compatible with an affected sibling. At birth, hematopoietic stem cells from the cord blood of the savior baby are then used to transplant the sick sibling. HLA typing alone is carried out for acquired diseases, such

as leukemias and anemias, or it can be combined with the detection of mutations underlying immunodeficiencies and hemoglobinopathies (Kuliev et al., 2005; Tur-Kaspa and Jeelani, 2015; Kakourou et al., 2014).

Although PGD is available for numerous nuclear monogenic disorders, it has rarely been used to test for mitochondrial (mt) DNA mutations (Bredenoord et al., 2008, 2009; Smeets et al., 2015). It is another ethically difficult indication, as PGD for mtDNA mutations reduces the risk for an affected child rather than eliminating it. Moreover, PGD can only reliably be offered in cases where the mtDNA mutation load closely relates with the disease severity and does not vary over time. Another criterion is a constant mutation load in all embryonic cells to ensure that results from a single cell biopsied will be predictive for the entire embryo.

In principle, any monogenic disease for which the chromosomal locus is unequivocally identified can be diagnosed with PGD. Molecular genetic reports stating the disease-causing mutation or at least the disease-associated phase within the family are mandatory at the intake of a PGD request, as this information will largely determine the strategy for genetic testing. Amplification methods are required to increase the limited amount of available genomic DNA (gDNA) in the embryonic samples. Targeted PCR amplification of the disease-associated locus has been applied since the start of PGD, whereas WGA methods were introduced more recently. Amplification is carried out following embryo biopsy: the single or multiple biopsied cells are washed, transferred to reaction tubes, and lysed. Amplification reaction components are then added directly to the lysed cell(s) without prior DNA purification. The methods for genetic testing have been continuously refined, and this has led to high levels of efficiency and accuracy. However, the amplification of a single or a few cells remains technically challenging because of the small amount of nonpurified DNA at the start and the inherent pitfalls of contamination, preferential amplification, and ADO, which may lead to misdiagnosis. As a high number of amplification cycles are applied to increase the low amount of DNA, contamination with extraneous DNA or carryover from previous amplification reactions is a major problem. High working standards and rigorous prevention measures should be adapted to prevent and control for contamination. ADO is defined as the random amplification failure of one of two alleles in a heterozygous cell sample, while in preferential amplification, one allele is less amplified than the other. Both are related problems because an under-amplified allele can drop out when detection methods of low sensitivity are used. The use of optimized cell lysis and DNA amplification conditions along with sensitive detection systems should reduce ADO to a minimum. Contamination and ADO measures have been described in the ESHRE PGD Consortium best practice guidelines for amplification-based PGD (Harton et al., 2011c).

23.3.2 Targeted Polymerase Chain Reaction-Based Amplification

Fluorescent multiplex PCR has been the gold standard in PGD for monogenic diseases for a long time. The method refers to the coamplification of several target sequences in one PCR reaction, using primer pairs in which one primer is fluorescently labeled, allowing for the subsequent sizing of PCR fragments on an automated sequencer or genome analyzer. In direct testing, the detection of the specific mutation(s) is combined with the amplification of multiple short tandem repeat (STR) markers, while in indirect or linkage-based testing only STR markers are coamplified (Laurie et al., 2010). The accuracy of the test is increased by using these closely linked markers, as they not only allow for indirect mutation detection, but they also allow for the detection of ADO, contamination, and recombination. The main advantage of linkage-based testing over mutation-specific testing is that the tests are mutation-independent, which makes them useful for several couples because of the polymorphic nature of the STR markers. This saves time, resources, and manpower in pre-PGD workups, especially in the case of diseases for which many private mutations have been identified. Whether the mutation amplicon is incorporated in the single cell test depends on multiple factors, including the nature of the mutation (familial or de novo), the mutation frequency, the availability of relevant family DNA samples, and the mutation type. The direct detection of complex and/or larger gene rearrangements is often not feasible, as single cell PCR fragments usually remain below 500 bp. The deletion or duplication mutations of a few nucleotides can be detected directly via fragment length differences, but single nucleotide changes can only be identified in post-PCR reactions. Over the years, different strategies of PCR and allele discrimination have been developed for single nucleotide mutation detection. The most important ones are the amplification refractory mutation system (Kokkali et al., 2007), endonuclease restriction (Moutou et al., 2007), minisequencing (Spits et al., 2006a), and quantitative real-time PCR (Bermudez et al., 2003). Examples of pre-PGD workup results (A) and PGD results (B) carried out via an indirect targeted multiplex PCR are shown in Fig. 23.3.

The initial step during the development of a single cell multiplex PCR test involves the in silico selection of STR markers, relying on public databases (University of California Santa Cruz, National Center for Biotechnology Information, Ensembl), followed by a primer design through software tools such as Primer3 together with the Basic Local Alignment Search Tool and SNPCheck searches to ensure the specificity of STR and mutation amplicons. Subsequently, STR marker genotyping is performed on DNA samples of the couple and family members with a known genetic status to identify informative markers and to establish which alleles of the informative markers segregate with the mutation. Ideally, a robust single cell multiplex PCR contains two to four informative markers located in both the upstream and downstream 1-Mb flanking

Indication	OMIM disease	Inheritance	Gene	OMIM gene	Ref Seq	Mat mutation	Pat mutation
Charcot-Marie-Tooth disease, type 1A	118220	AD	*PMP22*	601097	NM_000304.2	gene duplication	/

(A)

Informativity results						
Locus	Location relative to gene	Informativity	Family members			
			aff female	male	father of female	aff mother
Amelogenin			116 - 116	116 - 121	116 - 116	116 - 121
PMP22						
D17S2230	210 Kb 5'	Informative	268 - (278-294)*	257 - 288	263 - 268	294 - (278-294)*
D17S122	41 Kb 5'	Informative	216 - (214-216)*	206 - 214	208 - 216	216 - (214-216)*
PMP22str23AC	15 Kb 3'	Informative	164 - (152-164)*	162 - 162	152 - 164	152 - 152-164)*
D17S2229	74 Kb 3'	Not Informative	231 - (231-231)*	225 - 233	239 - 231	241 - (231-231)*
D17S2227	517 Kb 3'	Informative	177 - (152-167)*	162 - 157	157 - 177	152 - (152-167)*
D17S2220	790 Kb 3'	Informative	316 - (322-323)*	314 - 318	297 - 316	310 - (322-323)*

(B)

			Tubing	OK	OK	OK	OK	OK	OK	OK	OK
aff female	male	Locus	Relative location	E1.1	E3.1	E5.1	E6.1	E7.1	E8.1	E10.1	E11.1
PMP22											
268 - (278-294)*	257 - 288	D17S2230 FAM	210 Kb 5'	268 - 257	268 - 257	268 - 288	(278-294)* - 288	268 - 257	(278-294)* - 288	(278-294)* - 288	268 - 257
216 - (214-216.)*	206 - 214	D17S122OTN HEX	41 Kb 5'	216 - 206	216 - 206	216 - 214	(214-216.)* - 214	216 - 206	(214-216.)* - 214	(214-216.)* - 214	216 - 206
164 - (152-164)*	162 - 162	PMP22str23AC NED	15 Kb 3'	164 - 162	164 - 162	164 - 162	(152-164)* - 162	164 - 162	(152-164)* - 162	(152-164)* - 162	164 - 162
231 - (231-231)*	225 - 233	D17S2229 NED	74 Kb 3'	231 - 225	231 - 225	231 - 233	(231-231)* - 233	231 - 225	(231-231)* - 233	(231-231)* - 233	231 - 225
177 - (152-167)*	162 - 157	D17S2227 FAM	517 Kb 3'	177 - 162	177 - 162	177 - 157	(152-167)* - 157	177 - 162	(152-167)* - 157	(152-167)* - 157	177 - 162
316 - (322-323)*	314 - 318	D17S2220 HEX	790 Kb 3'	316 - 314	316 - 314?	316 - 318	(322-323)* - 318	316 - 314	(322-323)* - 318	(322-323)* - 318	316 - 314
Negative Control Intern: OK			Blanks	ok	ok	ok	ok	ok	ok	ok	ok
Negative Control Extern:			Remarks								
			DIAGNOSIS CELL	Unaffected	Unaffected	Unaffected	Affected	Unaffected	Affected	Affected	Unaffected
			DIAGNOSIS EMBRYO	Unaffected	Unaffected	Unaffected	Affected	Unaffected	Affected	Affected	Unaffected

FIGURE 23.3 Single cell targeted multiplex polymerase chain reaction informativity results (A) and embryo results (B) for a couple at risk for autosomal dominant Charcot-Marie-Tooth disease type 1A, caused by a duplication encompassing the *PMP22* gene. During the pre-preimplantation genetic diagnosis (PGD) workup, six short tandem repeat (STR) markers located in the duplication and flanking *PMP22* were tested on genomic DNA samples of the affected female and male partner; five STR markers were informative. Haplotypes for the affected female were determined using genomic DNA samples of her affected mother and unaffected father (phasing of some STR alleles, for instance, PMP22str23AC, was only finalized after the first PGD cycle). Haplotypes of the male were determined in the first PGD cycle. During the clinical cycle one cell was biopsied on day 3. Results of the PGD cycle (B) showed five unaffected and three affected embryos.

region of the mutation site. Primer sets are first tested and multiplexed on gDNA with the adaptation of reaction components and conditions before further fine-tuning at the single cell level. The optimized single cell protocol is then validated on a series of single cells and corresponding blanks, and only protocols yielding high-amplification efficiencies (preferably above 95%) and low contamination and ADO rates (both preferably below 5%) are accepted for clinical PGD application. Single or multiple lymphocytes, lymphoblasts, fibroblasts, or buccal cells are used as cell models to mimic either single blastomere or TE cell samples. As this process of single cell test adaptation and validation has to be repeated for every new DNA locus, it represents a major bottleneck for targeted PCR-based methods.

The method of PCR-based haplotyping with prior single cell WGA followed by standard PCR reactions of a multitude of STRs flanking the mutation site is a more generic method than targeted single cell PCR because adaptation/validation of PCR reactions to the single cell level can be omitted from the workup (Renwick et al., 2010).

23.3.3 Whole-Genome Amplification

WGA has been introduced as a universal step prior to karyotyping or genotyping to overcome the problem of the minute quantity of gDNA available for preimplantation testing and the long optimization time for the development of targeted PCR-based tests. The aim of WGA is to amplify the entire genome of single (or a few) cells up to several micrograms, which is sufficient for several standard downstream applications. The first WGA methods were PCR-based and suffered from incomplete genome coverage and amplification bias. The use of *Taq* DNA polymerase yielded an average fragment length of 400–500 bp (with a maximum size of 3 kb) and introduced many DNA sequence errors.

The multiple displacement amplification (MDA) method relying on isothermal strand displacement amplification has been widely used in single cell applications (Coskun and Alsmadi, 2007). In an MDA reaction, random exonuclease-resistant primers anneal to the denatured target DNA, and a DNA polymerase with strand–displacement

activity such as Phi29 elongates the primers in an isothermal reaction at 30°C. Additional priming events can occur on each displaced strand, leading to a network of branched DNA strands up to 10 kb. Because of the proofreading activity of the Phi29 polymerase, the error rate of MDA-based WGA is much lower compared with *Taq* DNA polymerase-based methods. A main disadvantage of MDA is the high rate of ADO and preferential amplification (25% on average) (Spits et al., 2006b).

A new generation of WGA methods combines MDA and PCR amplification. The SurePlex protocol (marketed as PicoPlex by Rubicon Genomics) initiates with DNA fragmentation and a preamplification MDA reaction using hybrid primers, followed by PCR (Langmore, 2002). The Multiple Annealing and Looping-Based Amplification Cycles (MALBAC) method uses specific primers that contain a common 27-nucleotide sequence and eight variable nucleotides for random annealing (Zong et al., 2012). The initial five MDA-based cycles generate hairpin amplicons because of the complementary ends of the MALBAC primers, precluding the use of these amplicons as a template in subsequent cycles and promoting the amplification of only original gDNA. The hairpin amplicons are subsequently amplified exponentially through PCR using primers complementary to the common region of the MALBAC primers.

None of the WGA methods produces a true linear representation of the single cell genome, and the results vary in ADO, preferential amplification rate, coverage, and nucleotide copy errors. As a consequence, a specific WGA method is chosen in the function of the downstream application: SurePlex is the method of choice for the detection of chromosomal copy numbers with array-comparative genomic hybridization (CGH) or next-generation sequencing (NGS) because of reduced amplification bias, and MDA is preferred for genotyping applications of PCR-based haplotyping or single nucleotide polymorphism (SNP) arrays in the case of monogenic disorders because of its better genome coverage and low error rates (Handyside et al., 2004; Deleye et al., 2015; de Bourcy et al., 2014).

23.4 PREIMPLANTATION GENETIC DIAGNOSIS FOR CHROMOSOMAL ABERRATIONS

23.4.1 Indications

It is estimated that 1 in 625 (0.16%) individuals carries a balanced chromosomal rearrangement, with translocation (reciprocal or Robertsonian) and inversions as the most frequent. Reciprocal translocations are structural chromosomal aberrations that result from the breakage of nonhomologous chromosomes. The broken segments are exchanged, forming derivative chromosomes. In most cases the translocation takes place between two nonhomologous chromosomes, but more complex rearrangements involving more chromosomes also occur. An enormous number of reciprocal translocations have been described, with breakpoints scattered throughout all chromosomal regions. With few exceptions, the majority of reciprocal translocations are family-specific. Robertsonian translocations originate through centric fusion of the long arms of any two of the five acrocentric chromosomes (De Vos et al., 2009; Scott et al., 2013; Gardner and Lane, 1998; Woodson et al., 2014; Derks-Smeets et al., 2014), with loss of the short arms. An inversion is a two-break event involving just one chromosome. The segment between the breaks is inverted and reinserted, and the breaks reunite.

Carriers of confirmed balanced chromosomal rearrangement are phenotypically normal. The production of unbalanced gametes, due to unfavorable meiotic segregation when the derivative chromosomes pair up at meiosis I, puts them at risk for repeated miscarriages, sub-or infertility, and unbalanced offspring. This explains why the incidence of balanced chromosomal rearrangement rises to about 3% in couples with reproductive problems. The theoretical chance of producing normal or balanced gametes is 4 out of 32 for reciprocal translocation carriers and 4 out of 16 for Robertsonian translocation carriers. However, the actual percentage depends on several factors, including the chromosomes involved, the location of the breakpoints, and the gender of the carrier. For carriers of an inversion, the chromosomal imbalance occurs if there is, within the inverted segment, a crossover at the first meiotic division between the inversion chromosome and the normal homolog, resulting in a recombinant chromosome, which may produce four different segregates.

23.4.2 Fluorescence In Situ Hybridization-Based Analysis

FISH (see also Chapter 14) utilizes chromosome-specific DNA probes labeled with different fluorochromes that are hybridized to the target DNA (interphase or metaphase) fixed to a microscopic slide. Microscopic evaluation and computerized imaging systems enable fluorescent probe signals to be identified and counted. As a result, FISH detects a chromosome or a part of a chromosome, thereby allowing for the determination of the copy number of that region in a particular sample.

Initial applications using FISH for diagnosing chromosomal translocations involved painting probes for metaphase chromosomes of PBs. A major shortcoming of this method was that only translocations of the female could be examined. Subsequently, homemade probes, isolated from cosmid (artificial plasmids containing the cos sequences of lambda phage), yeast artificial chromosome, or bacterial artificial chromosome (BAC) libraries, spanning or flanking the translocation breakpoints were applied on the interphase nucleus of a blastomere from cleavage-stage embryos. These probes gave very accurate

results because they distinguished between normal and balanced genotypes, but they were very time-consuming to isolate (Munné et al., 2000).

The commercial availability of centromeric and subtelomeric probes opened completely new perspectives for PGD for translocation carriers. PGD for Robertsonian translocations is relatively simple and involves a dual-color FISH with a single probe for each chromosome, with the exception of chromosomes 13 and 21 for which two probes per chromosome have been recommended, yielding a three-color FISH. The situation is more complex for reciprocal translocations. Ideally the protocol should include a probe for both centric segments and both subtelomeric regions of the translocated segments. In this case two scoring errors would be required to misdiagnose an unbalanced product as normal/balanced. However, four suitable probes may not be available. Most centers performing PGD for reciprocal translocations use a combination of two subtelomeric probes distal to the breakpoint, along with a centromeric probe for one of the two chromosomes involved to support the distinction between normal/balanced genotypes and unbalanced genotypes (Scriven et al., 1998).

Carriers of reciprocal translocations and inversions usually have unique rearrangements, necessitating the development, preclinical testing, and optimization of individualized probe mixtures for each couple. The limited number of distinct fluorochromes available for labeling DNA probes restricts the number of chromosomes that can be assessed in a single analysis, although PGD for structural chromosomal abnormalities usually interrogates only the chromosomes involved in the translocation. Single cell FISH is technically difficult; the initial step of cell fixation is critical and requires skill and experience. Probes may fail to hybridize, and the accuracy of interpretation can be affected by signal overlap or split signals.

23.4.3 Polymerase Chain Reaction-Based Analysis

PCR-based protocols, which utilize informative STR markers on both segments (flanking the breakpoint) of the translocated chromosomes, have also been applied in PGD for translocations (Fiorentino et al., 2010). Similar to the FISH-based testing, the assessment is targeted, and the analysis is usually limited to the chromosomes involved in the translocation. PCR-based protocols overcome several inherent limitations of FISH-based tests, as they are not dependent on cell fixation nor on microscopic signal interpretation. Moreover, the inheritance of individual chromosomes can be assessed and allows for the detection of uniparental disomy. Again, workup is necessary prior to any clinical case in order to identify informative STR markers.

23.4.4 Comparative Genomic Hybridization and Array-Comparative Genomic Hybridization-Based Analysis

The technology for chromosome analysis in PGD has shifted toward methods allowing for the analysis of all 24 chromosomes. The first clinically applied method was CGH (see also Chapter 14). This method employs a competitive hybridization of differentially labeled DNA samples (a test sample DNA with green fluorochrome and a chromosomally normal reference DNA with red fluorochrome) to normal metaphase chromosomes on a microscopic slide. The ratio of green versus red fluorescence along the length of each chromosome is evaluated with image-processing software, revealing the relative number of chromosome copies in the test sample compared with the reference. An excess of green fluorescence on a specific chromosome is indicative of a chromosomal gain, whereas an excess of red fluorescence is indicative of chromosome loss. WGA (see Section 23.3.3) is an obligatory step in the adaptation of CGH to the single cell level. The obvious advantage of CGH over FISH is that the copy number of all 24 chromosomes can be determined. In addition, CGH provides a more detailed picture of the entire length of each chromosome, enabling the detection of imbalanced chromosomal segments. CGH was successfully applied in PGD cycles for aneuploidy screening (Wilton et al., 2001) and a shortened protocol (12 h instead of 2−3 days) in PGD for translocations (Rius et al., 2011). Nevertheless, the difficult interpretation of results and low resolution of 10−20 Mb (Malmgren et al., 2002) have hampered the widespread use of CGH in PGD, and the method has largely been replaced by array-CGH.

Array-CGH is based on the same principle as CGH, but the metaphase spreads on the microscope slide have been replaced with genomic clones or DNA probes (microarray). Each probe is specific to a different chromosomal region and occupies a discrete spot on the slide. An analysis is performed by scanning and imaging the array and measuring the intensity of both hybridization signals relative to each probe (the logR ratio). Similar to CGH, an amplification step is required to increase the initial DNA quantity from a single or few biopsied cells. As array-CGH has a lower hybridization time than CGH it can be performed within the time frame of a PGD cycle with a fresh blastocyst transfer. Other advantages over CGH are the higher resolution and automation.

Different types of single cells, including PB, blastomeres, and TE cells, have been successfully tested for chromosome imbalances such as aneuploidies, unbalanced translocations, deletions, and duplications using array-CGH on different platforms. The platform most commonly used (24sure, Illumina, San Diego, CA, USA) contains about 3000 DNA fragments of 100−200 kb derived from loci

1 Mb apart across the genome, cloned in BAC; this so-called BAC array yields a resolution of about 20 Mb. A higher resolution array with an increased number of clones, particularly in the telomeres of each chromosome, is available (24sure+, Illumina, San Diego, CA, USA) for the detection of chromosomal imbalances in embryos of translocation carriers involving segments of about 5−10 Mb. A paired comparison between array-CGH and PCR-based testing on 200 embryos from translocation carriers was performed as a validation study of the high-resolution 24sure + BAC array. A high percentage of embryos (93%) were successfully diagnosed with array-CGH, and the study demonstrated that array-CGH analysis can identify segmental aneusomies as small as 2.5 Mb (Fiorentino et al., 2011). Apart from errors directly related to the chromosomes involved in the translocation, it has been postulated that the segregation patterns could have an influence on meiotic synapsis and the disjunction of other bivalents, causing aneuploidy in chromosomes not involved in the translocation. The existence of this so-called interchromosomal effect is controversial, and conflicting results have been reported in the literature (Gianaroli et al., 2002). Apart from the detection of specific chromosome imbalances in embryos from translocation carriers, array-CGH offers the benefit of concurrent aneuploidy screening for all 24 chromosomes, making it possible to study possible interchromosomal effects or other postzygotic chromosome errors.

Array-CGH provides a generic platform, avoiding the need to develop locus- and family-specific tests. The vast majority of balanced translocations may be reliably diagnosed by microarray. This applies to all Robertsonian translocations and the substantial majority of reciprocal translocations.

23.5 EMERGING TECHNOLOGIES

Various genome-wide platforms have become available for PGD: array-CGH, SNP array, and NGS (see also Chapters 14, 18, 19). These technologies all require WGA prior to application, and the final data interpretation has to take into account bias and artifacts introduced from WGA. Artifacts from the cell cycle phase should also be considered. It has been shown that the accuracy of chromosomal imbalances detected by BAC array-CGH is reduced for single S-phase cells (Dimitriadou et al., 2014). An analysis of TE samples with cells in different cell cycle stages may overcome this problem, while the WGA representation bias may be partially filtered out by computational algorithms but cannot be completely eliminated.

The different genome-wide platforms vary in resolution and the type of information they provide. Array-CGH can reveal chromosome imbalances, but polyploidies, balanced translocations, or inversions will not be detected. In contrast, SNP arrays, which enable chromosomal constitution evaluation in addition to genotyping, can detect copy neutral loss of heterozygosity or uniparental disomy. A limitation of SNP array is that de novo mutations cannot be detected, making the method unsuitable in the case of monogenic disorders, which frequently involve de novo mutations. NGS technologies have emerged; they offer simultaneously genotype and chromosome copy number data with increased accuracy, reliability, and resolution, presenting a universal platform for the evaluation of monogenic disorders (including the detection of de novo mutations), aneuploidy screening, and intra- and interchromosomal unbalanced structural variants.

SNP arrays are high-density oligo arrays containing up to several million probes of smaller length, which allow for the genotyping of hundreds of thousands of selected SNPs across all chromosomes in a single reaction. SNPs are mostly biallelic; the most common allele is indicated as A, and the less common allele is indicated as B. The commercially available SNP arrays use different chemistries for SNP genotyping. One method, for instance, is based on multiple SNP allele-specific probes to which fragmented and labeled test DNA will hybridize. Another method relies on SNP locus-specific probes that end one nucleotide before the SNP and to which fragmented test DNA hybridizes, and probes undergo single base extension reactions with differently labeled nucleotides (LaFramboise, 2009). The arrays are scanned, and SNP genotypes are called based on the ratio of hybridization intensities for A and B (allele frequencies) (e.g., AB is called in the case of similar intensities of an intermediate level). SNP array analysis is more detailed, as genotype calls are combined with copy number data. Two measures provide information about the copy number state: the logR ratio (the log2 transformed value of the normalized intensity of the SNP) and the B allele frequency (BAF, which is the signal intensity of the B allele over the total signal intensity for an SNP). BAF values of 0, 0.5, and 1 represent a normal copy number ($n = 2$), but aberrations will cause a decrease or increase of the total intensity and allele frequencies. SNP arrays at the level of a single cell yield a lot of noise because of WGA pitfalls and therefore demand particularly well-developed algorithms for data interpretation. Different SNP genotyping and different SNP copy number typing algorithms are available. Handyside and coworkers developed karyomapping, a family-based computational phasing approach for the reconstruction of haplotypes of SNPs flanking mutations (Handyside et al., 2010; Natesan et al., 2014). Karyomapping uses discrete diploid SNP calls (assuming AA, BB, AB, or No call as possible states for each SNP) together with basic Mendelian laws and requires a close relative for phasing. The underlying copy number state of

each SNP is not really considered. Genotyping algorithms using discrete diploid SNP calls will yield errors across regions with copy number variations (true or WGA-induced). The method of karyomapping is therefore quite suitable for the diagnosis of monogenic disorders but has restrictions in the detection of copy number aberrations. A novel method, called haplarithmisis, overcomes this limitation: it primarily relies on continuous BAFs and allows for haplotype and copy number detection as well as the determination of the parental origin of the chromosomal anomaly (Zamani Esteki et al., 2015). Haplarithmisis is a family-based approach necessitating samples of the couple and a close relative, and it is applicable for both SNP arrays and NGS. Examples of pre-PGD workup results and PGD results carried out via karyomapping with an SNP array are shown in Fig. 23.4.

NGS involves DNA fragmentation and the preparation of a library of templates using adapters that may contain barcodes for a more affordable analysis with multiple samples in a single run. The single molecule templates are then sequenced in parallel from one end or from both ends, either directly or after prior clonal amplification, and the sequence reads are mapped to a reference genome. A crucial parameter is the genome coverage or read depth, referring to the number of times a particular read was found at a given genomic position. This read depth is proportional to the copy number, and a relatively low coverage of 0.07x has been demonstrated as sufficient for accurate numerical and structural chromosome analysis by NGS (Yin et al., 2013). The same group presented a clinical validation study showing no differences in clinical outcome when comparing NGS to the SNP array, although NGS was able to detect some segmental imbalances more precisely (Tan et al., 2014).

For monogenic disorders, a targeted NGS-based approach with an increased read depth across the mutation site (minimum 100x) has been reported (Treff et al., 2013). The mutation sites were captured by amplification with mutation-specific primers (TaqMan allelic discrimination assay) prior to sequencing. A primer pool for chromosome copy number analysis was added at the amplification step for parallel aneuploidy screening via real-time quantitative PCR.

Further evaluation and thorough validation are required before NGS can be part of clinical PGD practice. As NGS technologies undergo constant improvement, costs get reduced, and the current aim is high throughput for postnatal and cancer applications, it may take some time before workflows, computational pipelines, and instruments have been adapted to the flexible needs of PGD services.

The implementation of array and NGS technologies will reduce the PGD workup substantially and introduce automation. Conversely, these technologies will generate a tremendous amount of genetic information, more than was initially requested, and it is expected that this will entail many ethical discussions and challenges for genetic counseling.

23.6 CLINICAL OUTCOME OF PREIMPLANTATION GENETIC DIAGNOSIS

The ESHRE PGD consortium has collected data on all of the details of PGD/PGS cycles, pregnancies, deliveries, and children annually since 1997. An overview has been presented after 10 years of data collection (Harper et al., 2012).

With respect to PGD for monogenic disorders, the latest report (cumulative data I-XIII) summarizes data from 9267 cycles, yielding a clinical pregnancy rate of 24% per oocyte retrieval (OR) and 30% per ET (De Rycke et al., 2015). Compared to other indications, PGD for monogenic disorders showed the highest pregnancy rate. The rate is lower than can be expected in a regular IVF cycle, but it must be taken into account that a large cohort of embryos is diagnosed as abnormal or affected (50% and 25% chance in the case of autosomal dominant and autosomal recessive disorders, respectively). The genetic selection is strong, especially in patients requesting HLA compatibility, with a 25% chance to identify an HLA-matched embryo in the case of HLA alone or a 19% chance in the case of HLA combined with an autosomal recessive disorder. A small set of embryos has no successful diagnosis; the percentage of embryos without diagnosis has gradually decreased from 19% in the early data to 9% in the latest data sets, mainly because of continuous technical improvements.

PGD for chromosomal aberrations has shown that large proportions of the embryos are chromosomally imbalanced and are not genetically suitable for transfer (an average of 75% versus 57% in monogenic disorder cases). As the highest levels of chromosomally abnormal embryos are found in the reciprocal translocation group, fewer embryos are available for transfer, and lower pregnancy rates are achieved compared with the Robertsonian translocation group (16% per OR and 27% per ET for reciprocal translocation cases versus 23% per OR and 31% per ET for Robertsonian translocation cases). The data collections showed that PGD for reciprocal translocations was performed more often than for any other type of structural chromosome abnormality (58%). This has an impact on the global outcome for PGD for chromosomal aberrations, with cumulative data I-XIII demonstrating an overall pregnancy rate of 18% per OR and 28% per ET procedure. It is clear that when embryos are available for transfer, the chances for pregnancy are similar as for other indications (28% per ET versus 29% per ET for monogenic disorder cases).

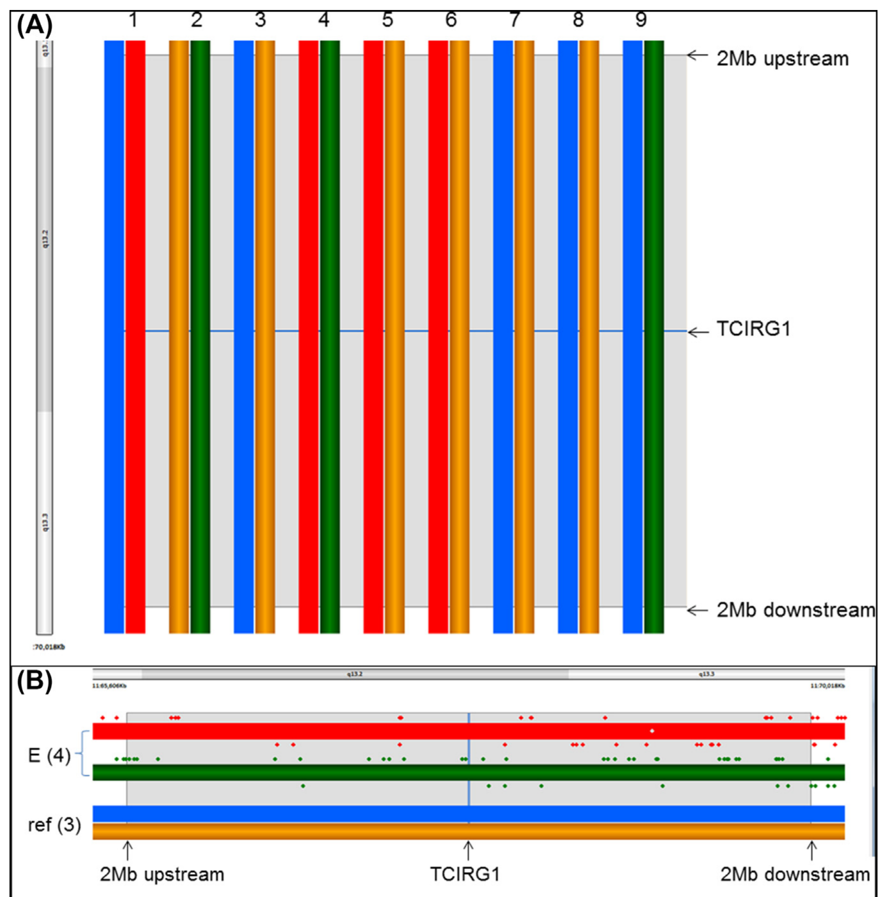


FIGURE 23.4 Single nucleotide polymorphism (SNP) array karyomapping case view (A) and detailed karyomapping (B) results for a couple at risk for autosomal recessive Osteopetrosis, caused by mutations in the *TCIRG1* gene on chromosome 11 (67,806,483−67,818,362 (+) assembly GRCh37/hg19). During the pre-preimplantation genetic diagnosis (PGD) workup, genomic DNA samples of the carrier father (1), carrier mother (2), and an affected son (3) were analyzed to determine the number of informative SNPs in the region of the *TCIRG* gene (informative SNP => one parent is homozygous, and the other parent is heterozygous at the SNP locus). During the PGD cycle, these results are combined with embryonic samples (4−9). The case view (A) shows the paternal mutant haplotype in blue (*Dark gray in print versions*) and the paternal wild type haplotype in red (*Gray in print versions*); the maternal mutant haplotype is shown in orange (*Light Gray in print versions*) and the maternal wild type haplotype in green (*Darkest gray in print versions*). The blue (*Dark gray in print versions*) horizontal line represents the gene; the surrounding zone in lightest gray represents a 2-Mb region upstream and downstream of the gene. Embryo (4) is unaffected, embryos (5 and 6) are carriers of the maternal mutation. Embryos (7 and 8) are affected, and embryo (9) is a carrier of the paternal mutation. Detailed karyomapping result (B) for embryo (4) (top), relative phasing of SNPs versus the reference (bottom) showing the *TCIRG1* gene (blue (*Dark gray in print versions*) vertical line) and 2-Mb flanking regions (*lightest gray zone*). Blue (*Dark gray in print versions*) and orange (*Light Gray in print versions*) represent parental haplotypes inherited by the reference; red (*Gray in print versions*) and green (*Darkest gray in print versions*) represent parental haplotypes not inherited by the reference. The *dots* on top of the chromosomes indicate "key"SNPs; the *dots* below the chromosomes indicate "non-key"SNPs; and the *white dots* in the middle of the chromosomes indicate SNPs that were not called. The presence of 25 maternal and 13 paternal key SNPs within the 2-Mb flanking region supports the presence of wild type parental haplotypes, diagnosing embryo (4) as unaffected, homozygous wild type. ("Key"SNPs are SNP genotypes of the sample at an informative SNP locus, which cannot be altered by allele dropout; "non-key"SNPs are SNP genotypes that are most likely accurate, but errors cannot be excluded due to allele dropout. Example: father AB, mother AA => informative SNP; sample BA => key SNP; sample AA => non-key SNP).

The cumulative PGD consortium data sets I-XIII on 8453 children born after PGD as well as multiple follow-up studies on children born after PGD show that pregnancies and children are very similar to pregnancies and children born after ICSI. Embryo biopsy, mostly at the cleavage stage, is the major technical difference between PGD and ICSI. One report on a prospective cohort of 242 PGD children (Eldar-Geva et al., 2014) and another large series of 995 children (Desmyttere et al., 2012) demonstrated no adverse neonatal outcomes and no additional risk of embryo biopsy. Developmental neurological and cognitive assessment and follow up on psychomotor and social functioning showed that PGD preschoolers were comparable with controls born after ICSI or after spontaneous conception (Winter et al., 2014, 2015; Sacks et al., 2015). Further monitoring on the safety of the procedures is necessary, as the long-term health outcome remains unknown.

23.7 ACCURACY AND QUALITY CONTROL

ADO and contamination are inherent pitfalls of amplification-based preimplantation testing, while signal overlap, hybridization failure, and nonspecificity constitute technical difficulties during FISH procedures. Apart from these technical errors, possible causes for misdiagnosis involve intrinsic sample quality, like chromosomal mosaicism, and human errors, such as mislabeling or incorrect ET.

The true misdiagnosis rate is difficult to assess, as many ETs have no follow up (yield no pregnancy or birth), only a minority of children have pre-or postnatal testing, and just a small number of nontransferred and non-cryopreserved embryos are reanalyzed. Although centers are encouraged to report misdiagnosis cases to the PGD consortium (anonymously) so that other centers may learn from them, the accounts of a 0.14% misdiagnosis rate for PCR-based cycles and 0.06% for FISH-based PGD cycles are most likely underestimations (Wilton et al., 2009). For PCR-based cycles, one multicenter retrospective study based on the reanalysis of nearly 1000 embryos showed a high diagnostic accuracy of 93.7% (Dreesen et al., 2014). PGD does not assure the proper constitution of the future embryo but minimizes the risk of the disorder under investigation.

Reanalysis of embryos donated by the couples is an important parameter in quality control and quality assurance. External quality assessment (EQA) is another key element of quality assurance (Vendrell et al., 2009; Harper et al., 2010). The available EQA schemes for monogenic diseases and for FISH- or array-based testing were initiated by the Cytogenetic External Quality Assessment Service and the United Kingdom National External Quality Assessment Services in collaboration with the ESHRE PGD consortium. Participation in EQA schemes yields an independent evaluation of a laboratory's methods, results interpretation, and reporting and allows for a comparison against other PGD centers. The yearly EQA scheme reports also contain valuable information that helps laboratories to improve on their techniques and reports.

Accreditation or assessment against international standards (the general quality management standard ISO9001 and ISO15189, which is specific for medical laboratories) by an independent accreditation body has been recommended and even been required as an integral part of genetic testing in many countries. In 2008, only 33% of 53 European PGD centers had achieved or were preparing for accreditation (Corveleyn et al., 2008). Results from a PGD consortium survey in 2014 showed an improvement of about half of 46 IVF centers and half of 46 diagnostic laboratories in the field of PGD having obtained accreditation, according to national or international standards (unpublished data).

23.8 CONCLUSIONS AND FUTURE PERSPECTIVES

PGD evolved from an experimental procedure in the early 1990s to a widely accepted alternative for prenatal diagnosis in its first decade. PGD was applied for an expanding range of indications, and technical improvements provided reliable and accurate single cell genetic tests for chromosomal abnormalities and monogenic disorders.

More advancements have been introduced in the fields of preimplantation genetic testing and ARTs; the major innovations are a switch from targeted to genome-wide testing and from cleavage-stage to blastocyst biopsy as well as the introduction of vitrification as a successful embryo cryopreservation method. The applications of powerful array and NGS technologies following universal WGA reduce the workload and allow for standardization and uniformity in the complex process of PGD. The high amplification needed in the case of a single or a few cells as starting materials introduces amplification artifacts (ADO, errors, bias, and chimeric products), and this complicates downstream data interpretation. As this is different from genome-wide haplotyping and copy number analysis of unamplified samples, a new generation of bioinformatic tools is required, enabling genome-wide analysis at the single cell level, bearing in mind the artifacts.

As innovative algorithms are being developed, and the cost of sequencing continues to decline, whole-genome sequencing in the context of PGD will become available in the near future. This will generate a tremendous amount of genetic data, for instance, on disease susceptibility genes, entailing challenges for genetic counseling and raising new ethical issues. It is clear that the rapid technological advances in the field of PGD should be balanced with ethical reflection and thorough discussions.

WEB RESOURCES

The URLs for the data presented are as follows:

Preimplantation Genetic Diagnosis International Society: http://www.pgdis.org
European Society of Human Reproduction and Embryology Preimplantation Genetic Diagnosis Consortium:
http://www.eshre.eu/Data-collection-and-research/Consortia/PGD-Consortium.aspx
University of California Santa Cruz: http://genome.ucsc.edu/
National Center for Biotechnology Information: http://www.ncbi.nlm.nih.gov/
Ensembl: http://www.ensembl.org
Primer3: http://primer3.ut.ee/
Basic Local Alignment Search Tool: http://blast.ncbi.nlm.nih.gov/Blast.cgi

United Kingdom National External Quality Assessment Services:
http://www.ukneqas-molgen.org.uk
Cytogenetic External Quality Assessment Service: http://www.ceqas.org/

COMPETING INTERESTS

The authors declare that they have no competing interests.

REFERENCES

Bermudez, M.G., Piyamongkol, W., Tomaz, S., Dudman, E., Sherlock, J.K., Wells, D., 2003. Single-cell sequencing and mini-sequencing for preimplantation genetic diagnosis. Prenat. Diagn. 23, 669–677.

Bredenoord, A.L., Dondorp, W., Pennings, G., De Die-Smulders, C.E., De Wert, G., 2008. PGD to reduce reproductive risk: the case of mitochondrial DNA disorders. Hum. Reprod. 23, 2392–2401.

Bredenoord, A., Dondorp, W., Pennings, G., de Die-Smulders, C., Smeets, B., de Wert, G., 2009. Preimplantation genetic diagnosis for mitochondrial DNA disorders: ethical guidance for clinical practice. Eur. J. Hum. Genet. 17, 1550–1559.

Checa, M.A., Alonso-Coello, P., Solà, I., Robles, A., Carreras, R., Balasch, J., 2009. IVF/ICSI with or without preimplantation genetic screening for aneuploidy in couples without genetic disorders: a systematic review and meta-analysis. J. Assist. Reprod. Genet. 26, 273–283.

Corveleyn, A., Morris, M.A., Dequeker, E., et al., 2008. Provision and quality assurance of preimplantation genetic diagnosis in Europe. Eur. J. Hum. Genet. 16, 290–299.

Coskun, S., Alsmadi, O., 2007. Whole genome amplification from a single cell: a new era for preimplantation genetic diagnosis. Prenat. Diagn. 27, 297–302.

de Bourcy, C.F., De Vlaminck, I., Kanbar, J.N., Wang, J., Gawad, C., Quake, S.R., 2014. A quantitative comparison of single-cell whole genome amplification methods. PLoS One 9, e105585.

De Rycke, M., Belva, F., Goossens, V., et al., 2015. ESHRE PGD Consortium data collection XIII: cycles from January to December 2010 with pregnancy follow-up to October 2011. Hum. Reprod. 30, 1763–1789.

De Vos, A., Van Steirteghem, A., 2001. Aspects of biopsy procedures prior to preimplantation genetic diagnosis. Prenat. Diagn. 21, 767–780.

De Vos, A., Staessen, C., De Rycke, M., et al., 2009. Impact of cleavage-stage embryo biopsy in view of PGD on human blastocyst implantation: a prospective cohort of single embryo transfers. Hum. Reprod. 24, 2988–2996.

Deleye, L., Coninck, D.D., Christodoulou, C., et al., 2015. Whole genome amplification with SurePlex results in better copy number alteration detection using sequencing data compared to the MALBAC method. Sci. Rep. 5, 11711.

Derks-Smeets, I.A., de Die-Smulders, C.E., Mackens, S., et al., 2014. Hereditary breast and ovarian cancer and reproduction: an observational study on the suitability of preimplantation genetic diagnosis for both asymptomatic carriers and breast cancer survivors. Breast Cancer Res. Treat. 145, 673–681.

Desmyttere, S., De Rycke, M., Staessen, C., et al., 2012. Neonatal follow-up of 995 consecutively born children after embryo biopsy for PGD. Hum. Reprod. 27, 288–293.

Dimitriadou, E., Van der Aa, N., Cheng, J., Voet, T., Vermeesch, J.R., 2014. Single cell segmental aneuploidy detection is compromised by S phase. Mol. Cytogenet. 7, 46.

Dokras, A., Sargent, I.L., Ross, C., Gardner, R.L., Barlow, D.H., 1990. Trophectoderm biopsy in human blastocysts. Hum. Reprod. 5, 821–825.

Dreesen, J., Destouni, A., Kourlaba, G., et al., 2014. Evaluation of PCR-based preimplantation genetic diagnosis applied to monogenic diseases: a collaborative ESHRE PGD consortium study. Eur. J. Hum. Genet. 2, 1012–1018.

Eldar-Geva, T., Srebnik, N., Altarescu, G., et al., 2014. Neonatal outcome after preimplantation genetic diagnosis. Fertil. Steril. 102, 1016–1021.

Fiorentino, F., Kokkali, G., Biricik, A., et al., 2010. Polymerase chain reaction-based detection of chromosomal imbalances on embryos: the evolution of preimplantation genetic diagnosis for chromosomal translocations. Fertil. Steril. 94, 2001–2011.

Fiorentino, F., Spizzichino, L., Bono, S., et al., 2011. PGD for reciprocal and Robertsonian translocations using array comparative genomic hybridization. Hum. Reprod. 26, 1925–1935.

Gardner, D.K., Lane, M., 1998. Culture of viable human blastocysts in defined sequential serum-free media. Hum. Reprod. 13 (Suppl. 3), 148–159.

Gianaroli, L., Magli, M.C., Ferraretti, A.P., et al., 2002. Possible interchromosomal effect in embryos generated by gametes from translocation carriers. Hum. Reprod. 17, 3201–3207.

Gianaroli, L., Magli, M.C., Pomante, A., et al., 2014. Blastocentesis: a source of DNA for preimplantation genetic testing. Results from a pilot study. Fertil. Steril. 102, 1692–1699.

Handyside, A.H., Kontogianni, E.H., Hardy, K., Winston, R.M., 1990. Pregnancies from biopsied human preimplantation embryos sexed by Y-specific DNA amplification. Nature 344, 768–770.

Handyside, A.H., Robinson, M.D., Simpson, R.J., et al., 2004. Isothermal whole genome amplification from single and small numbers of cells: a new era for preimplantation genetic diagnosis of inherited disease. Mol. Hum. Reprod. 10, 767–772.

Handyside, A.H., Harton, G.L., Mariani, B., et al., 2010. Karyomapping: a universal method for genome wide analysis of genetic disease based on mapping crossovers between parental haplotypes. J. Med. Genet. 47, 651–658.

Harper, J., Coonen, E., De Rycke, M., et al., 2010. What next for preimplantation genetic screening (PGS)? A position statement from the ESHRE PGD Consortium Steering Committee. Hum. Reprod. 25, 821–823.

Harper, J.C., Sengupta, S., Vesela, K., et al., 2010. Accreditation of the PGD laboratory. Hum. Reprod. 25, 1051–1065.

Harper, J.C., Wilton, L., Traeger-Synodinos, J., et al., 2012. The eshre PGD consortium: ten years of data collection. Running title: 10 years of PGD data Hum. Reprod. Update 18, 234–247.

Harton, G., Braude, P., Lashwood, A., , et al.European Society for Human Reproduction and Embryology (ESHRE) PGD Consortium, 2011a. ESHRE PGD consortium best practice guidelines for organization of a PGD centre for PGD/preimplantation genetic screening. Hum. Reprod. 26, 14–24.

Harton, G.L., Harper, J.C., Coonen, E., Pehlivan, T., Vesela, K., Wilton, L., European Society for Human Reproduction and Embryology (ESHRE) PGD Consortium, 2011b. ESHRE PGD consortium best practice guidelines for fluorescence in situ hybridization-based PGD. Hum. Reprod. 26, 25–32.

Harton, G.L., De Rycke, M., Fiorentino, F., et al., European Society for Human Reproduction and Embryology (ESHRE) PGD Consortium, 2011c. ESHRE PGD consortium best practice guidelines for amplification-based PGD. Hum. Reprod. 26, 33–40.

Harton, G.L., Magli, M.C., Lundin, K., Montag, M., Lemmen, J., Harper, J.C., European Society for Human Reproduction and Embryology (ESHRE) PGD consortium/Embryology special interest group, 2011d. ESHRE PGD consortium/Embryology special interest group—best practice guidelines for polar body and embryo biopsy for preimplantation genetic diagnosis/screening (PGD/PGS). Hum. Reprod. 26, 41–46.

Kakourou, G., Destouni, A., Vrettou, C., Traeger-Synodinos, J., Kanavakis, E., 2014. A generic, flexible protocol for preimplantation human leukocyte antigen typing alone or in combination with a monogenic disease, for rapid case work-up and application. Hemoglobin 38, 49–55.

Kokkali, G., Traeger-Synodinos, J., Vrettou, C., et al., 2007. Blastocyst biopsy versus cleavage stage biopsy and blastocyst transfer for preimplantation genetic diagnosis of beta-thalassaemia: a pilot study. Hum. Reprod. 22, 1443–1449.

Kuliev, A., Rechitsky, S., Verlinsky, O., et al., 2005. Preimplantation diagnosis and HLA typing for haemoglobin disorders. Reprod. Biomed. Online 11, 362–370.

LaFramboise, T., 2009. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. Nucleic Acids Res. 37, 4181–4193.

Langmore, J.P., 2002. Rubicon Genomics, Inc. Pharmacogenomics 3, 557–560.

Laurie, A.D., Hill, A.M., Harraway, J.R., et al., 2010. Preimplantation genetic diagnosis for hemophilia A using indirect linkage analysis and direct genotyping approaches. J. Thromb. Haemost. 8, 783–789.

Malmgren, H., Sahlén, S., Inzunza, J., et al., 2002. Single cell CGH analysis reveals a high degree of mosaicism in human embryos from patients with balanced structural chromosome aberrations. Mol. Hum. Reprod. 8, 502–510.

McArthur, S.J., Leigh, D., Marshall, J.T., de Boer, K.A., Jansen, R.P., 2005. Pregnancies and live births after trophectoderm biopsy and preimplantation genetic testing of human blastocysts. Fertil. Steril. 84, 1628–1636.

Moutou, C., Gardes, N., Nicod, J.C., Viville, S., 2007. Strategies and outcomes of PGD of familial adenomatous polyposis. Mol. Hum. Reprod. 13, 95–101.

Munné, S., Sandalinas, M., Escudero, T., Fung, J., Gianaroli, L., Cohen, J., 2000. Outcome of preimplantation genetic diagnosis of translocations. Fertil. Steril. 73, 1209–1218.

Natesan, S.A., Bladon, A.J., Coskun, S., 2014. Genome-wide karyomapping accurately identifies the inheritance of single-gene defects in human preimplantation embryos in vitro. Genet. Med. 16, 838–845.

Preimplantation Genetic Diagnosis International Society (PGDIS), 2008. Guidelines for good practice in PGD: programme requirements and laboratory quality assurance. Reprod. Biomed. Online 16, 134–147.

Renwick, P., Trussler, J., Lashwood, A., Braude, P., Ogilvie, C.M., 2010. Preimplantation genetic haplotyping: 127 diagnostic cycles demonstrating a robust, efficient alternative to direct mutation testing on single cells. Reprod. Biomed. Online 20, 470–476.

Rius, M., Obradors, A., Daina, G., et al., 2011. Detection of unbalanced chromosome segregations in preimplantation genetic diagnosis of translocations by short comparative genomic hibridization. Fertil. Steril. 96, 134–142.

Roque, M., Lattes, K., Serra, S., et al., 2013. Fresh embryo transfer versus frozen embryo transfer in in vitro fertilization cycles: a systematic review and meta-analysis. Fertil. Steril. 99, 156–162.

Sacks, G.C., Altarescu, G., Guedalia, J., et al., 2015. Developmental neuropsychological assessment of 4- to 5-year-old children born following Preimplantation Genetic Diagnosis (PGD): a pilot study. Child. Neuropsychol. 1–14.

Scott, R.T., Upham, K.M., Forman, E.J., Zhao, T., Treff, N.R., 2013. Cleavage-stage biopsy significantly impairs human embryonic implantation potential while blastocyst biopsy does not: a randomized and paired clinical trial. Fertil. Steril. 100, 624–630.

Scriven, P.N., Handyside, A.H., Ogilvie, C.M., 1998. Chromosome translocations: segregation modes and strategies for preimplantation genetic diagnosis. Prenat. Diagn. 18, 1437–1449.

Smeets, H.J., Sallevelt, S.C., Dreesen, J.C., de Die-Smulders, C.E., de Coo, I.F., 2015. Preventing the transmission of mitochondrial DNA disorders using prenatal or preimplantation genetic diagnosis. Ann. NY Acad. Sci. 1350, 29–36.

Spits, C., De Rycke, M., Verpoest, W., Lissens, W., Van Steirteghem, A., Liebaers, I., Sermon, K., 2006a. Preimplantation genetic diagnosis for Marfan syndrome. Fertil. Steril. 86, 310–320.

Spits, C., Le Caignec, C., De Rycke, M., et al., 2006b. Whole-genome multiple displacement amplification from single cells. Nat. Protoc. 1, 1965–1970.

Tan, Y., Yin, X., Zhang, S., et al., 2014. Clinical outcome of preimplantation genetic diagnosis and screening using next generation sequencing. Gigascience 3, 30.

Tobler, K.J., Zhao, Y., Ross, R., et al., 2015. Blastocoel fluid from differentiated blastocysts harbors embryonic geneomic material capable of a whole-genome deoxyribonucleic acid amplification and comprehensive chromosome micro-array analysis. Fertil. Steril. 104, 418–425.

Treff, N.R., Scott Jr., R.T., Su, J., et al., 2012. Polar body morphology is not predictive of its cell division origin. J. Assist. Reprod. Genet. 29, 137–139.

Treff, N.R., Fedick, A., Tao, X., Devkota, B., Taylor, D., Scott Jr., R.T., 2013. Evaluation of targeted next-generation sequencing-based preimplantation genetic diagnosis of monogenic disease. Fertil. Steril. 99, 1377–1384.

Tur-Kaspa, I., Jeelani, R., 2015. Clinical guidelines for IVF with PGD for HLA matching. Reprod. Biomed. Online 30, 115–119.

Vendrell, X., Carrero, R., Alberola, T., et al., 2009. Quality management system in PGD/PGS: now is the time. J. Assist. Reprod. Genet. 26, 197–204.

Verlinsky, Y., Kuliev, A., 1996. Preimplantation diagnosis of common aneuploidies in fertile couples of advanced maternal age. Hum. Reprod. 11, 2076–2077.

Wilton, L., Williamson, R., McBain, J., Edgar, D., Voullaire, L., 2001. Birth of a healthy infant after preimplantation confirmation of

euploidy by comparative genomic hybridisation. N. Engl. J. Med. 34, 1537–1541.

Wilton, L., Thornhill, A., Traeger-Synodinos, J., Sermon, K.D., Harper, J.C., 2009. The causes of misdiagnosis and adverse outcomes in PGD. Hum. Reprod. 24, 1221–1228.

Winter, C., Van Acker, F., Bonduelle, M., Desmyttere, S., De Schrijver, F., Nekkebroeck, J., 2014. Cognitive and psychomotor development of 5- to 6-year-old singletons born after PGD: a prospective case-controlled matched study. Hum. Reprod. 29, 1968–1977.

Winter, C., Van Acker, F., Bonduelle, M., Desmyttere, S., Nekkebroeck, J., 2015. Psychosocial development of full term singletons, born after preimplantation genetic diagnosis (PGD) at preschool age and family functioning: a prospective case-controlled study and multi-informant approach. Hum. Reprod. 30, 1122–1136.

Woodson, A.H., Muse, K.I., Lin, H., et al., 2014. Breast cancer, BRCA mutations, and attitudes regarding pregnancy and preimplantation genetic diagnosis. Oncologist 19, 797–804.

Xu, K., Montag, M., 2012. New perspectives on embryo biopsy: not how, but when and why? Semin. Reprod. Med. 30, 259–266.

Yin, X., Tan, K., Vajta, G., et al., 2013. Massively parallel sequencing for chromosomal abnormality testing in trophectoderm cells of human blastocysts. Biol. Reprod. 88, 69.

Zamani Esteki, M., Dimitriadou, E., Mateiu, L., 2015. Concurrent whole-genome haplotyping and copy-number profiling of single cells. Am. J. Hum. Genet. 96, 894–912.

Zong, C., Lu, S., Chapman, A.R., Xie, X.S., 2012. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science 338, 1622–1626.

Chapter 24

Noninvasive Cell-Free DNA Prenatal Testing for Fetal Aneuploidy in Maternal Blood

V. Velissariou[1] and P.C. Patsalis[2]
[1]Bioiatriki Health Services, Athens, Greece; [2]The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus

24.1 INTRODUCTION

Noninvasive prenatal testing (NIPT) of the fetal genotype in maternal plasma without the risk of miscarriage of invasive procedures was the "holy grail" of prenatal diagnosis for many years (Bianchi, 2004). Today, NIPT for aneuploidy using cell-free fetal DNA (cffDNA) in maternal plasma is a reality, and it is revolutionizing prenatal screening and diagnosis. The potential aim is to avoid invasive procedures such as chorionic villus and amniotic fluid sampling, which result in a significant pregnancy loss risk between 0.5% and 1% (Tabor and Alfiveric, 2010). The discovery of cffDNA circulating in the maternal blood (Lo et al., 1997) has provided scientists with a great opportunity for the development of NIPT methodologies. Such strategies have been successfully applied for the determination of the fetal rhesus status and inherited monogenic disease in cases where the fetus does not carry the maternal mutation (Tounta et al., 2011). However, the field of fetal aneuploidy investigation seems to be more challenging. The main reason for this is that the maternal cell-free DNA (cfDNA) in the mother's plasma is far more abundant and is identical to half of the corresponding fetal DNA (Lo et al., 1997; Ashoor et al., 2013). Methodologies developed are mainly based on next-generation sequencing (NGS) technology and epigenetic genetic modifications, such as fetal−maternal DNA differential methylation (Papageorgiou et al., 2014). Current NIPT methodologies, based on counting DNA sequences using NGS, involve whole-genome sequencing (WGS), targeted sequencing, and the assessment of single nucleotide polymorphism (SNP) differences between the mother and fetus (Lo et al., 2013; Lau et al., 2014). Such methods are employed worldwide and are the most rapidly adopted genomic tests. Clinical trials have demonstrated the efficacy of NIPT for trisomy 21 (Down syndrome), trisomy 18 (Edwards syndrome), and possibly for trisomy 13 (Patau syndrome) and the sex chromosome abnormalities (Gil et al., 2015). Investigators have also described the NIPT of subchromosomal imbalances, opening up further clinical possibilities (Wapner et al., 2015). Moreover, developments in the field have allowed for noninvasive prenatal determination of fetal methylomic and transcriptomic profiles (Lo et al., 2013; Tsui et al., 2014), making NIPT one of the most exciting and fast-developing fields in both research and the clinical setting.

24.2 ESTABLISHED PRENATAL SCREENING AND DIAGNOSIS PRACTICES

24.2.1 Screening

It is important to distinguish prenatal diagnosis of aneuploidy from antenatal screening. A diagnostic test performed by invasive procedures [chorionic villous sampling (CVS), amniocentesis, or fetal blood sampling] needs to be accurate since the result will be decisive as to whether to terminate the pregnancy. A false negative (FN) result (aneuploid fetus misdiagnosed as euploid) would lead to the birth of an abnormal child. On the other hand, a false positive (FP) result (euploid fetus misdiagnosed as aneuploid) could lead to the termination of a normal pregnancy. In contrast, antenatal screening does not aim to be

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00024-9
Copyright © 2017 Elsevier Ltd. All rights reserved.

definitive; rather, it is designed to identify women who are at sufficiently high risk of common aneuploidies as to warrant invasive prenatal diagnosis. Since the invasive diagnostic procedures, mainly CVS and amniocentesis, are hazardous and expensive, only a relatively small group of women are identified as being at high risk.

Initially, the determination of a high-risk population was based on maternal age alone, and hundreds of invasive procedures were performed to find one affected fetus. The selection of women at risk has improved over the years, and established screening programs include biochemical screening and the evaluation of ultrasonographic markers, which determine the risk of Down syndrome (trisomy 21), Edwards syndrome (trisomy 18), Patau syndrome (trisomy 13) and Turner syndrome (45,X). The model-predicted aneuploidy detection rates (DRs) for fixed false−positive rates (FPRs) of established prenatal screening protocols have been extensively discussed in the literature (Benn et al., 2013).

24.2.2 Diagnosis

Chromosomal cytogenetic analysis of cultured chorionic villi and amniotic fluid cells can identify numerical chromosomal abnormalities and large structural rearrangements, mainly deletions, duplications, translocations, inversions, and extra supernumerary marker chromosomes. Since the culturing of chorionic villi and amniotic fluid cells requires at least 10−15 days, various methods have been developed, mainly in order to obtain a rapid result. These include fluorescence in situ hybridization (FISH) in uncultured amniotic cells (Morris et al., 1999) and quantitative fluorescence polymerase chain reaction (PCR) or multiplex ligation-dependent amplification in fetal DNA (Mann et al., 2012; Gerdes et al., 2008). Although these methods were designed to test for only the most frequent numerical chromosomal abnormalities, they are now widely used in routine clinical practice as an adjunct to full cytogenetic analysis (Caine et al., 2005). The most frequent numerical chromosomal abnormality in man is trisomy 21 (Down syndrome), with a birth rate of approximately 1 in 700 newborns (Gardner et al., 2012). Other frequent numerical chromosomal abnormalities are trisomy 18 (Edwards syndrome), trisomy 13 (Patau syndrome), and sex chromosome abnormalities, which result in Klinefelter (47,XXY) and Turner (45,X) syndromes (Gardner et al., 2012).

Fetal karyotyping, which results from conventional cytogenetic analysis by light microscopy, has been considered over the last 40 years to be the "gold standard" in prenatal diagnosis with very low error rates (Gardner et al., 2012; Benn, 2010). However, it is inefficient at detecting submicroscopic deletions and duplications, which are smaller than 5−10 million base pairs (5−10 Mb), which are often associated with malformations and mental retardation. About 15−20% of children with unexplained developmental delay/intellectual disability, autism spectrum disorders, or multiple congenital anomalies have subtle microdeletions or microduplications, which are not visible by light microscopy or conventional karyotyping (Sagoo et al., 2009).

Subtle abnormalities, which cannot be detected with the conventional G-banded karyotype, can be investigated and identified by chromosomal microarray analysis (CMA), with array comparative genomic hybridization (array-CGH) being the most widely used array-based methodology (Sagoo et al., 2009). Microdeletions and microduplications are submicroscopic copy number variants (CNVs) and are common in the human genome; however some can cause congenital malformations and syndromes (Cooper et al., 2011). Array-CGH has been routinely used in prenatal diagnosis, together with conventional fetal karyotyping, thus improving the detection of pathogenic fetal chromosomal imbalances (Evangelidou et al., 2013; Konialis and Pangalos, 2015) (Fig. 24.1). Analysis by CMA is highly recommended in fetuses with structural ultrasound abnormalities and a normal karyotype. It has been estimated that 3.1−9.1% of fetuses with structural ultrasound abnormalities and a normal karyotype will show a submicroscopic CNV of clinical significance (de Wit et al., 2014).

24.2.3 Hazards of Invasive Testing

The principal hazard of amniocentesis and CVS is miscarriage. A Cochrane Review (Alfirevic et al., 2003) showed that, when performed by a skilled operator, the fetal loss rates of the two procedures were comparable. An updated review (Tabor and Alfirevic, 2010) concluded that the excess miscarriage rate for either procedure is between 0.5% and 1%.

24.3 HISTORICAL BACKGROUND OF NONINVASIVE PRENATAL TESTING

24.3.1 Fetal Cells in the Maternal Circulation

It has been shown that various fetal cell types enter the maternal circulation during pregnancy (Bianchi, 2004). These include erythrocytes, platelets, trophoblast cells, leukocytes, and mesenchymal stem/progenitor cells. For many years, the focus of research efforts was on the physical separation and genetic analysis of these cells. About 1 in 10^3−10^7 nucleated cells in maternal blood are fetal. The proportion of fetal cells can be enriched to about 1 in 10−100 by techniques such as magnetic cell sorting or fluorescence-activated cell sorting after the attachment of cell surface markers. The resulting sample is unsuitable for traditional cytogenetic analysis because it is still highly

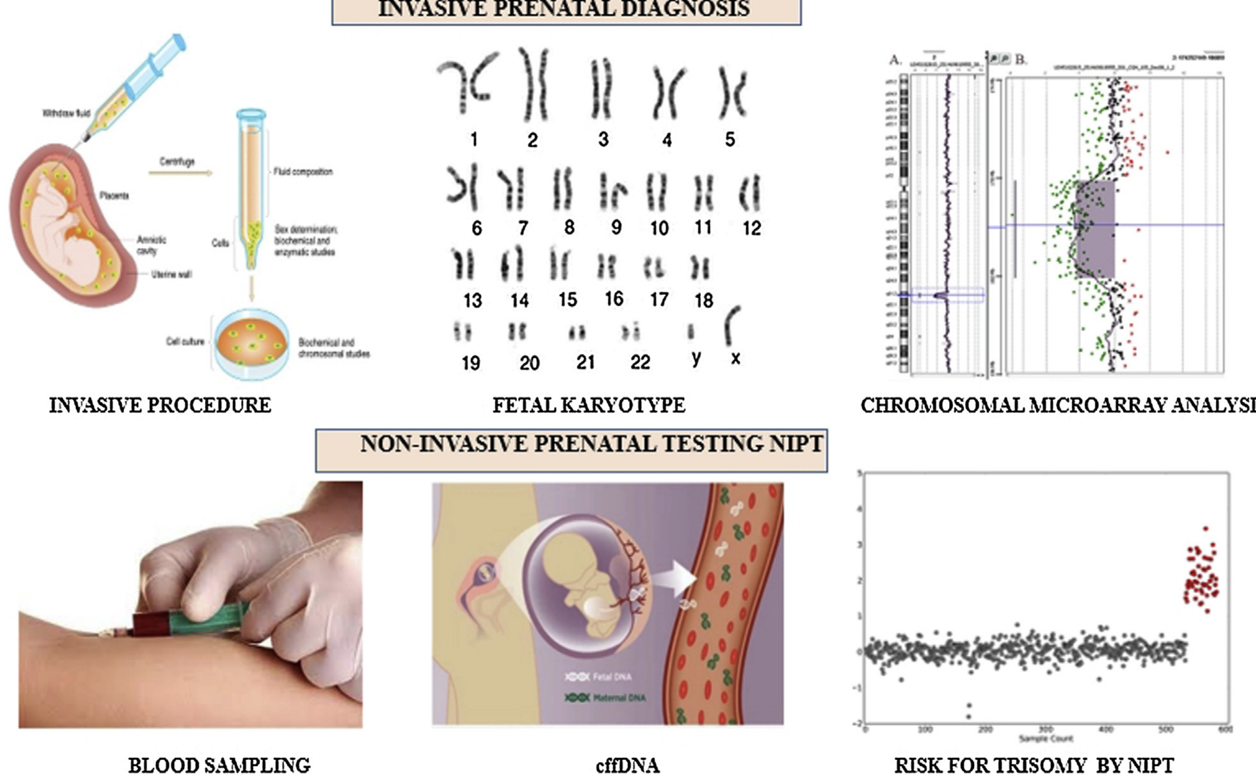


FIGURE 24.1 The traditional approach of fetal aneuploidy diagnosis by invasive procedures followed by cytogenetic analysis and/or chromosomal microarray analysis (top) and noninvasive prenatal testing (NIPT) for common fetal aneuploidies (bottom). *cffDNA*, cell-free fetal DNA.

contaminated with maternal cells. However, using chromosome-specific DNA probes and FISH, it is possible to suspect fetal trisomy by the presence of magnetically labeled or fluorescent antibodies on specific fetal three-signal nuclei in some of the cells of the maternal blood enriched for fetal cells. In particular, nucleated erythrocytes seemed to be good candidates for chromosomal aneuploidy testing using the FISH method (Bianchi et al., 2002). However, after many years of research, the international study sponsored by the National Institute of Child Health and Human Development (the "NIFTY Trial") concluded that while fetal cells are present in the maternal circulation, there was no method robust and reliable enough to isolate them for widespread clinical application (Bianchi et al., 2002).

24.3.2 Cell-Free Fetal DNA in the Maternal Circulation

An important advance in the area of NIPT was the discovery by Lo et al. (1997) that cffDNA circulates within the blood of pregnant women. It is likely derived from apoptotic cells and is amplifiable, allowing for multiple prenatal applications. Using PCR of Y-specific DNA sequences and agarose-gel electrophoresis, Lo et al. (1997) were able to detect fetal DNA in most plasma/serum samples of pregnant women bearing male fetuses. The development of real-time PCR techniques enabled these authors to use a quantitative approach to measure the concentration of fetal DNA in maternal plasma (Lo et al., 1998a). The fractional fetal DNA concentrations in maternal plasma were remarkably high, having a mean of some 3% in early pregnancy and 6% in late pregnancy, much higher than the corresponding figures for fetal nucleated cells that are present in maternal blood. Furthermore, a serial analysis on a cohort of pregnant women sampled at multiple times during pregnancy indicated that fetal DNA was present in maternal serum as early as the seventh week of gestation. All in all, these data indicated that fetal DNA in maternal plasma could be a valuable material for noninvasive prenatal diagnosis, much more robust than using fetal cells in maternal blood. Following this, it was first shown by the same authors that the prenatal detection of fetal *RHD* sequences in maternal plasma is feasible (Lo et al., 1998b). This test was the first clinical application of DNA-based noninvasive prenatal diagnosis, and it has been used routinely in clinical practice ever since.

Enormous efforts have been made by several research teams toward developing NIPT for fetal aneuploidy (Benn et al., 2013; Allyse et al., 2015; Gil et al., 2015). The

pioneering work of Dennis Lo and his colleagues has had an important impact (Lo, 2013). Down syndrome is caused by an increased dosage of genetic material on chromosome 21, most commonly caused by the presence of an extra copy of the chromosome. A trisomy 21 fetus with three copies of chromosome 21, rather than two copies, releases proportionally more chromosome 21 DNA into maternal plasma than does a euploid fetus. The concentration of cffDNA in the plasma of pregnant women carrying Down syndrome fetuses is higher than those carrying chromosomally normal fetuses (Lo et al., 1999a,b). Hence maternal plasma DNA analysis can be used as a screening test for fetal Down syndrome (Fig. 24.1). The challenge was that there is some overlap between the concentration ranges of the Down syndrome group and the normal group, and thus the diagnostic sensitivity and specificity of the test needed to be improved before it could become a valuable testing tool. Hence the key to the NIPT of Down syndrome was the development of methods that would allow one to accurately measure such a chromosome dosage increase. In 2007, two independent reports described an approach that can be used for detecting the small increase in chromosome 21-derived genetic material in maternal plasma in trisomy 21 pregnancies using digital PCR (Lo et al., 2007a; Fan et al., 2007). The investigators reasoned that the degree of the increase would be dependent on the fractional concentration of fetal DNA in maternal plasma. They also deduced that, as the fractional concentration of fetal DNA decreases, more DNA molecules would need to be counted. However, because digital PCR only amplifies DNA molecules in maternal plasma that have binding sites for the predetermined PCR primers, the majority of plasma DNA molecules, which do not have the primer binding sites, are not analyzed, and thus the diagnostic information contained in them is wasted.

24.3.3 Free-Fetal RNA in the Maternal Circulation

Another approach that appeared to be promising was testing free-fetal RNA (ffRNA). The presence of ffRNA in maternal plasma was established by Poon et al. (2002), who showed that Y chromosome-specific zinc finger protein (*ZFY*) mRNA could be found in the plasma of women carrying a male fetus. The identification of genes that are expressed by trophoblasts but not maternal cells should result in the presence of the corresponding cffRNA species in maternal plasma, which is free of contaminating maternal RNA with a similar sequence. If the cffRNA also carries polymorphisms that distinguish fetal alleles, the presence of aneuploidy could be deduced by the departure from the normal 1:1 ratio in the relative amount of the inherited alleles (Lo et al., 2007b). This approach to NIPT showed some success in the initial proof-of-principle studies, but unpublished clinical trial data were disappointing.

24.3.4 Fetal–Maternal Differentially Methylated Regions

The discovery of fetal–maternal differentially methylated regions has facilitated the development of NIPT strategies (Chim et al., 2008; Papageorgiou et al., 2009). The aim of DNA methylation-based methodologies was first to identify fetal-specific methylation markers that would allow for the discrimination of fetal DNA from the maternal DNA in maternal circulation and that have the potential to be developed into noninvasive prenatal diagnostic markers.

DNA methylation is an enzymatic chemical modification of the genome, which includes the addition of a methyl group to the carbon-5 position of the cytokines of CpG dinucleotides (Schubeler et al., 2015). The methylation pattern of the cell is reset during embryogenesis, and it is established early during development (Monk, 1990). After its establishment, the methylation pattern is inherited from one cell generation to the next. DNA methylation is a dynamic process and may change during the post-developmental stage (Reik et al., 2001). It is believed that 60% of tissue-specific differentially methylated regions are methylated in embryonic cells, while during the differentiation of embryonic tissues to adult tissues, they undergo demethylation (Schubeler et al., 2015). It has been shown that specific regions of the genome show a different methylation pattern in different tissues and at different stages of development. The earlier discussed findings provided convincing evidence that fetal DNA will present different methylation patterns from the methylation pattern of the maternal DNA (Poon et al., 2002).

The most promising methylation-based approach was methylated DNA immunoprecipitation (MeDIP) (Papageorgiou et al., 2014). The identification of a set of specific DNA sequences on chromosome 21 that are hypermethylated in placental cells allows for the selective enrichment of fetal cfDNA from maternal plasma. Following quantitative real-time PCR, there may be measurable differences in the amount of chromosome 21-derived DNA when fetal trisomy 21 is present, compared with euploid fetuses. Using the "epigenetic approach," an NIPT method based on MeDIP combined with quantitative PCR was developed (Papageorgiou et al., 2011, 2014; Patsalis et al., 2012; Tsaliki et al., 2012).

24.3.5 Next-Generation Sequencing of Maternal Plasma DNA

The use of NGS methodologies in NIPT has revolutionized the field. In 2008, two independent groups demonstrated

that NIPT of trisomy 21 could be achieved using massively parallel shotgun sequencing (MPSS), ushering in a new era of NIPT and opening new possibilities for the use of these methodologies in clinical practice (Chiu et al., 2008; Fan et al., 2008). A large-scale, multicenter validation study of this methodology demonstrated a diagnostic sensitivity of 100% and a specificity of 98% (Chiu et al., 2011). The results were replicated by other groups (Palomaki et al., 2011; Bianchi et al., 2012). This methodology was introduced into clinical service in 2012 and as of 2014 had been performed on 700,000 maternal plasma samples in 50 countries. Commercially available NIPT identifies aneuploidies in chromosomes 13, 16, 18, 21, 22, X, and Y, allowing for the detection of the most frequently observed chromosome aneuploidies (data updated as of 2014) (Agarwal et al., 2013).

24.4 ORIGIN OF CELL-FREE FETAL DNA

Circulating cfDNA in healthy adults derives mainly from hematopoietic cells undergoing apoptosis (Lui et al., 2002). The exact origin of cffDNA remains unknown. It is suggested that it is derived mainly from the placenta, as demonstrated by the very rapid clearance of fetal DNA from maternal blood following delivery, in contrast to the majority of fetal cells that can survive several weeks postpartum (Hahn et al., 2005). Additional evidence that cffDNA is of placental origin only is the fact that it is detectable even in the absence of an embryo (Alberry et al., 2007). Faas et al. (2012) described a case in which the plasma DNA and amniocentesis results were 45,X. Cytogenetic studies of the placenta showed that the mesenchymal core had a 46,XX karyotype, and the cytotrophoblast layer was 45,X. This study is often cited as the basis for the fact that cffDNA derives from cytotrophoblasts (Fig. 24.2). In addition to this evidence, the absence of circulating fetal DNA species in case reports concerning placental mosaicism confirms the origin of cffDNA being the placenta and not the fetal unit itself (Masuzaki et al., 2004; Hahn et al., 2005). Other studies have suggested that the major source of cffDNA is the cells of trophoblastic origin that are released from the syncytiotrophoblast in the form of syncytial knots. These cells undergo apoptosis, and the nucleic acids packed inside, including RNA and DNA, are released into the maternal circulation. In addition to apoptotic mechanisms taking place as a result of normal aging of the syncytiotrophoblast, accidental breakage or necrosis may also be one of the causes of the release of cell-free nucleic acids. The concept that apoptotic or aponecrotic pathways of the syncytiotrophoblast may alter the proportion of fetal DNA is gathering considerable interest (Litton et al., 2009).

It is well established that an increase in cffDNA levels can be used as a predictive marker for the early detection of pregnancy-related disorders, such as preeclampsia, intrauterine growth restriction, preterm labor, placenta previa, and hyperemesis gravidarum, although conflicting evidence suggests that cfDNA levels may increase during the early stages of pathological changes and may later decrease as the disease progresses (Sifakis et al., 2015). Despite this discrepancy, cffDNA levels clearly increase prior to the onset of the clinical symptoms of pregnancy-related complications. The mechanisms of action are believed to be a combination of apoptotic, aponecrotic, and inflammatory events that manifest during placental development. Further and more intensive research is required to elucidate the exact pathways that govern the underlying pathology of the release of cffDNA and pregnancy-related complications, although the results presented thus far indicate that cffDNA should be considered as a serious addition to the field of noninvasive prenatal screening and the early evaluation of severe pregnancy-related complications.

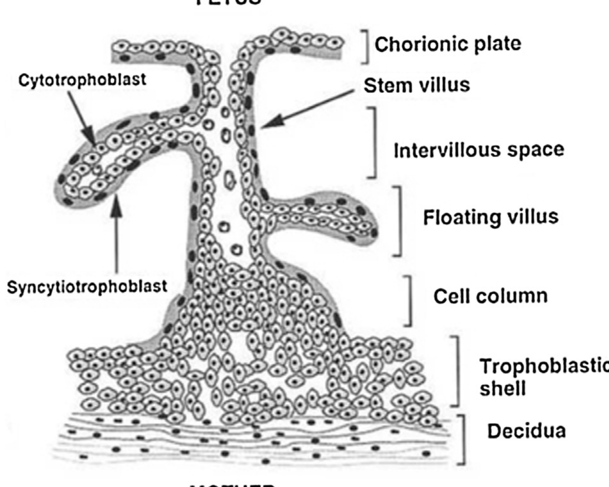


FIGURE 24.2 Schematic presentation of the chorionic villous.

24.5 NONINVASIVE PRENATAL TESTING METHODOLOGIES

The extraordinary advances in molecular biology and NGS have made NIPT possible. There are four main methodologies: WGS, SNP-based targeted sequencing, the targeted capture enrichment methodology, and the microarray-based targeted methodology. All methodologies analyze the total (maternal and fetal) cfDNA that is isolated from maternal plasma, which is fragmented. Fetal DNA fragments are predominantly 143 bp in length, and the maternal DNA fragments are generally 166 bp (Lo et al., 2010). The microarray-based targeted methodology combines

microarray methodology and digital analysis of selected regions (DANSR™) (Juneau et al., 2014).

In contrast to WGS, where the whole genome is sequenced, in the targeted methodologies only those chromosomal regions that are of interest (e.g., chromosomes 21, 18, and 13) are selectively amplified, and then it is assessed whether there is a departure from euploidy, based on the relative number of DNA fragment counts for this subset of chromosomes (Sparks et al., 2012; Koumbaris et al., 2016). These "targeted" methodologies have the advantage of requiring considerably less sequencing and thereby reducing costs. A minimum fetal fraction (*ff*), typically 4%, is required for all NIPT methodologies.

Most NIPT commercial providers use NGS-based tests. Each of them has different proprietary algorithms to determine the cutoffs between normal and abnormal results. The classification cutoff will determine the expected FPRs. Furthermore, certain companies include an adjustment to allow for the variability in the proportion of DNA that is fetal and then combines the results of the laboratory test with maternal age to provide a patient-specific risk for Down, Edwards, and Patau syndromes (Sparks et al., 2012). Commercial NIPT tests available worldwide are presented in Fig. 24.3.

24.5.1 Whole-Genome Sequencing

This approach relies on identifying and counting large numbers of the DNA fragments in the plasma specimen. Using WGS, millions of both fetal and maternal DNA fragments can be sequenced simultaneously, and since the entire human genome sequence is known, each piece that maps to a discrete locus can be assigned to the chromosome from which it came (Chiu et al., 2008; Fan et al., 2008) (Fig. 24.4).

Only the first 25 or 36 bp of each fragment are sequenced; this sequence is unique enough to allow alignment to a specific physical location within the human genome. A mapped sequence is known as a tag. The companies that perform WGS map between 12 and 25 million tags per sample. The number of tags at a specific chromosome is then counted and compared to the reference values obtained. If fetal aneuploidy is present, there should be a relative excess or deficit for the chromosome in question. The difference in counts will be small; for example, if fetal trisomy 21 is present, and the *ff* is 20%, the relative excess in chromosome 21 DNA fragments will be $(0.8 \times 2) + (0.2 \times 3) = 2.2$ compared with the situation for a euploid fetus $(0.8 \times 2) + (0.2 \times 2) = 2$, that is, a relative increase in the number of chromosome 21 counts of only 10%. To be able to reliably detect these differences, large numbers of counts are necessary, the *ff* needs to be appreciable, and the counts need to be compared with the expected counts for euploid cases. The latter can be achieved either by normalizing counts for the chromosome of interest against other chromosomes that are expected to be disomic within the same test run (Fan et al., 2008) or by comparing the fraction of counts assigned to a particular chromosome against the fractions seen for a set of known euploid cases (Sehnert et al., 2011). Results can be

Commercial NIPT tests available worldwide

- Whole Genome Sequencing (WGS)

- SNP-based Targeted Sequencing

- Targeted Capture Enrichment Technology

- Microarray-based Targeted

FIGURE 24.3 Noninvasive prenatal testing (NIPT) methods available worldwide showing the provider company and the methodology used. *SNP*, single nucleotide polymorphism.

FIGURE 24.4 Whole-genome sequencing methodology.

expressed as a Z-score, which can be converted into a probability that a specific chromosome result departs from the normal diploid situation. Although it is possible to convert the Z-score to a patient-specific risk, major commercial providers of this testing choose to present their results only in terms of "positive" or "negative" based on Z-scores exceeding predefined thresholds (Fig. 24.5).

A similar approach is known as "shotgun" WGS, which relies on sequencing and counting all informative chromosome regions (Fan et al., 2010; Dan et al., 2012; Liang et al., 2013). Sequences that do not map to a unique locus are uninformative. In some studies, only those fragments that have sequences identical to the reference genome sequence are accepted in the analysis, while in others one or two mismatches in the sequence will be accepted.

24.5.2 Single Nucleotide Polymorphism-Based Targeted Sequencing

It has long been shown that SNPs in the DNA are common among different individuals (Weiss, 1998; Kwok and Gu, 1999). The use of SNPs for noninvasive aneuploidy testing was first demonstrated by Dhallan et al. (2007). In their approach, the paternal, maternal, and fetal SNPs present in blood (paternal), buffy coat (maternal), and maternal plasma (maternal and fetal) were evaluated as bands on sequencing gels. The quantification of band intensities of the uniquely inherited paternal SNPs in the plasma allowed for an estimation of the *ff*. A comparison of the maternal plus fetal bands with the unique fetal band intensity also allowed for an estimation of the fetal chromosome 21 dosage (Fig. 24.6).

The SNP approach selectively amplifies and sequences 19,488 polymorphic loci on chromosomes 13, 18, 21, X, and Y (Zimmerman et al., 2012; Nikolaides et al., 2013). Each product is evaluated based on the hypothesis that the fetus is monosomic, disomic, or trisomic. After considering the positions of the SNPs on the chromosomes and the possibility that there may have been recombination, a maximum likelihood is calculated that the fetus is either normal, aneuploid (chromosome 21, 18, 13, or sex chromosome), triploid, or that uniparental disomy is present.

The algorithm used in this approach incorporates maternal genotype information and recombination frequencies to construct billions of theoretical fetal genotypes. The software calculates a relative likelihood for each hypothetical genotype by comparing it with data obtained from the maternal buffy coat DNA sequence. The SNP approach does not require reference values. It also avoids some of the technical problems associated with guanine cytosine (GC) bias and can detect fetal triploidy. However, because of the need to compare sequences with the maternal genotype, it cannot be used in egg donor pregnancies.

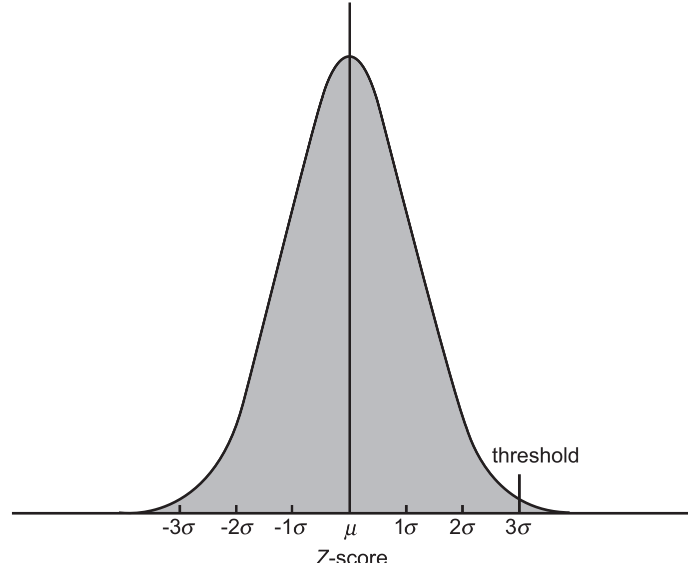


FIGURE 24.5 Standard normal distribution, which has a mean (μ) of zero and a standard deviation (σ) of 1, from which *Z*-scores are obtained for a widely used commercial noninvasive prenatal testing method. A typical threshold, also illustrated in the figure, would be set at *Z*-score = 3. This score corresponds to 3 standard deviations (3σ) away from the mean (μ) and would imply a large deviation from the normal/average population. As such, *Z*-score values greater than 3, which are found on the right-hand side of the threshold bar (*dark gray* portion of the distribution), would be considered "positive." Conversely, *Z*-scores of less than 3, which are found on the left-hand side of the threshold (light gray portion of the distribution) would be considered "negative." *Reproduced with permission from Koumbaris, G., Kypri E., Tsangaras K., Achilleos A., Mina P., Neofytou M., Velissariou V., Christopoulou G., Kallikas I., González-Liñán A., Benusiene E., Latos-Bielenska A., Marek P., Santana A., Nagy N., Széll M., Laudanski P., Papageorgiou E.A., Ioannides M., Patsalis P.C., 2016. Cell-free DNA analysis of targeted genomic regions in maternal plasma for non-invasive prenatal testing of trisomy 21, trisomy 18, trisomy 13 and fetal sex. Clin. Chem. 62, 848–855.*

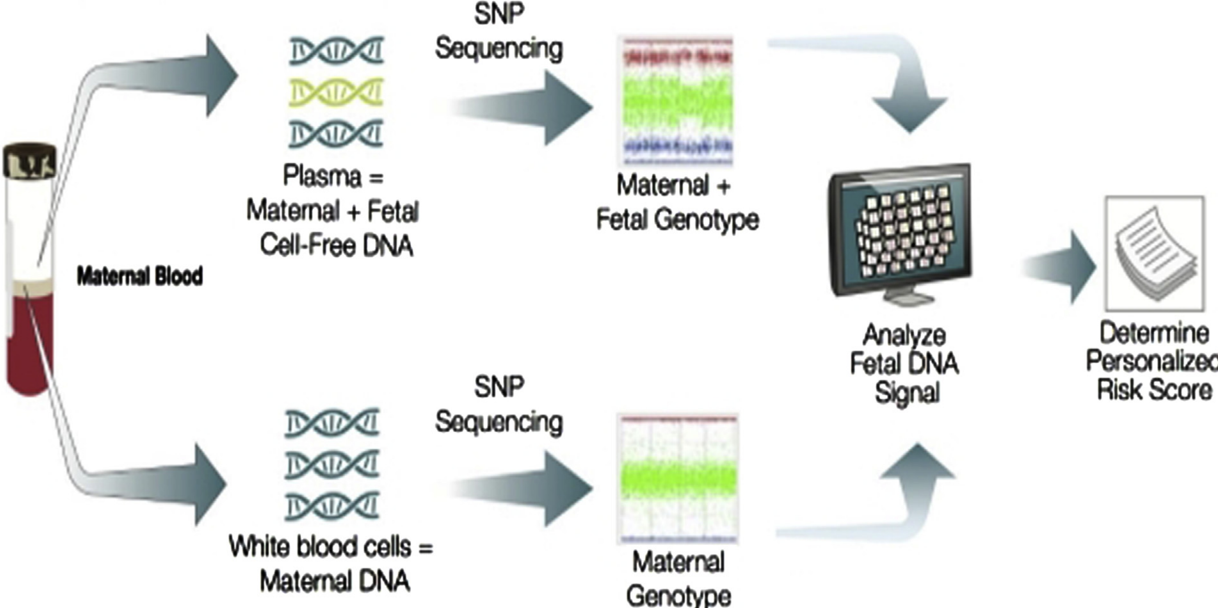


FIGURE 24.6 Single nucleotide polymorphism (SNP)-based targeting sequencing methodology.

In theory, with SNPs there should be advantages over methods based on total sequence counting. SNPs can provide information about the parent of origin of aneuploidy, recombination, and the inheritance of mutations. On the other hand, SNPs account for only 1.6% of the human genome, and therefore enrichment for fetal DNA, deeper sequencing, or higher levels of high-fidelity amplification may be required to identify unambiguously affected pregnancies with small imbalances (Liao et al., 2012). When few loci are used in the analysis, greater attention needs to be paid to the exclusion of regions in which there are rare benign CNVs. The SNP approach could result in the inadvertent detection of nonpaternity or perhaps consanguinity. In the case of assisted reproduction with an egg from an unrelated donor, SNP-based methodologies would need to be modified to take into consideration the presence of additional maternally derived fetal alleles not present in the surrogate mother.

24.5.3 Targeted Capture Enrichment Methodology

This targeted test constitutes an integrated assay that incorporates simultaneous determination of *ff* and accurate detection of fetal aneuploidies of chromosomes 13, 18, and 21 as well as fetal gender (Koumbaris et al., 2016) (Fig. 24.7). The assay employs a robust analysis algorithm, which minimizes random and systemic variation between sequencing runs and is sensitive enough to distinguish between euploid and aneuploid samples. There is a clear separation between the risk scores of trisomic and disomic samples, allowing for a binary classification scheme. The algorithm uses a Bayesian approach to estimate fetal DNA fraction. As such, additional information can be easily incorporated into the model. In addition, instead of inferring a point estimate of fetal DNA fraction, the algorithm calculates the posterior distribution of the fetal DNA fraction in each sample. It subsequently uses the lower bound of the corresponding 95% credible interval to determine whether a sample has adequate *ff*. This conservative approach of estimating *ff* ensures that the lowest possible *ff* of each sample is considered for classification purposes, thus minimizing the possibility of incorrect calls that could potentially arise from low proportions of fetal DNA. This novel *ff* estimation algorithm was also independently and thoroughly validated using Y chromosome loci in male samples. The *ff* estimation algorithm was also tested using nonpregnant samples. The algorithm correctly identified the absence of fetal DNA in these samples.

The targeted capture enrichment methodology has several advantages compared to WGS methods. WGS requires a very large number of reads and only allows for the simultaneous analysis of very few samples. The inherently limited throughput of whole genome methods imposes a significant technical, financial, and logistical burden. In contrast, the targeted method described here utilizes only specific genomic regions and significantly reduces the number of required reads. This results in a dramatic decrease in technical difficulties and a significant reduction in overall costs. Furthermore, the method permits very high read depth and provides very high accurate classification at a near-diagnostic level. The methodology also ensures very accurate *ff* estimation and protection from FN results. At the same time, the selection and enrichment of specific

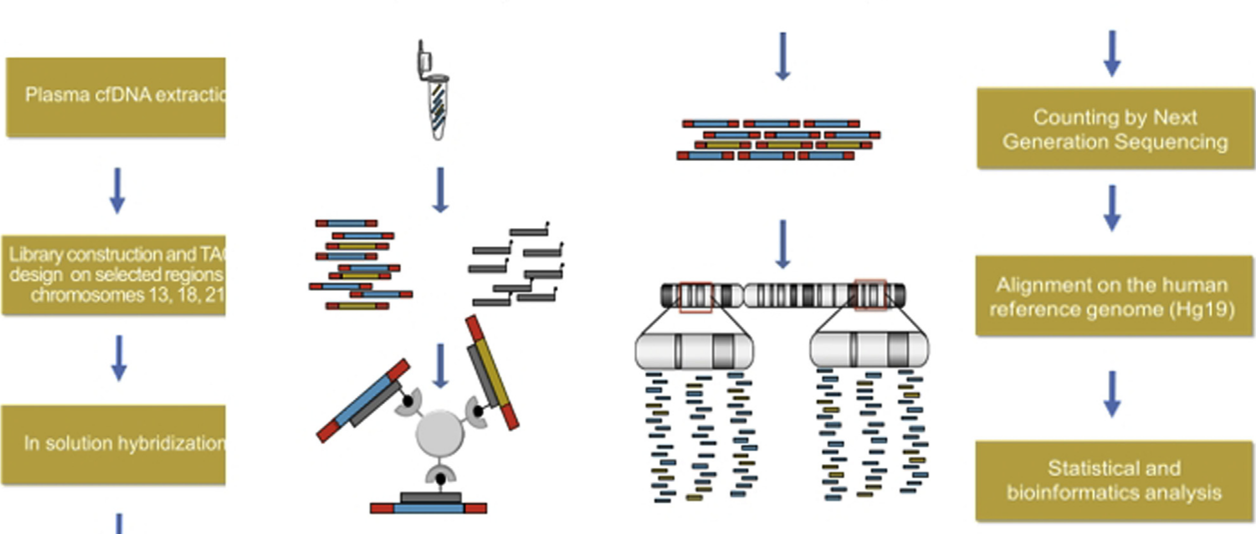


FIGURE 24.7 Targeted capture enrichment methodology. *cfDNA*, cell-free DNA.

targeted genomic regions of interest, called TACS, protects from FP and FN results because it avoids the phenomenon of GC-bias amplification as well as CNVs and regions of complex genomic architecture (Koumbaris et al., 2016).

24.5.4 Microarray-Based Targeted Methodology

Targeted chromosomes (13, 18, 21, X, and Y) are analyzed using DANSR assays. DANSR products are subsequently divided between two DNA quantification methods: microarrays and NGS. For both microarray and sequencing methodologies, the fetal-fraction optimized risk of trisomy evaluation algorithm is used to determine the trisomy risk and assay variability across samples and compute *ff* variability within samples (Juneau et al., 2014).

The DANSR targeted approach was developed to reduce the amount of sequencing required for NIPT. DANSR efficiently generates relevant data by focusing assay resources on chromosomes of clinical relevance.

Chromosomes 13, 18, and 21 together comprise only about 8% of the human genome. Targeting hundreds of DANSR assays to each of these chromosomes provides deep cfDNA analysis at a lower cost.

Microarray-based NIPT methodologies require no sample multiplexing. Instead, each sample is hybridized individually to a single microarray. The processing throughput is enhanced by physically connecting 384 microarrays onto a single multimicroarray plate for convenient high-throughput handling (Fig. 24.8). Because each sample is processed individually, and sample normalization is not required, time is saved and the cost is reduced.

24.6 BIOLOGICAL AND TECHNICAL FACTORS THAT AFFECT NONINVASIVE PRENATAL TESTING RESULTS

The accuracy of NIPT is affected by multiple technical and biological factors that are all integrated with and affected by each other. These include mainly *ff*, GC base content of the chromosomes tested, chromosome value, and maternal CNVs.

24.6.1 Fetal Fraction

The *ff* is the amount of cffDNA divided by the total DNA, which comprises fetal plus maternal cfDNA. The *ff* can be computed on the basis of polymorphic sequence variation (Norton et al., 2012) or methylation differences (Nygren et al., 2010) between maternal and fetal DNA. Most of the current serum markers used in Down syndrome screening follow a Gaussian curve. The correct measurement of the *ff* is of great importance in the accuracy of NIPT results. The higher the *ff*, the easier it is to call aneuploidy, because the *Z*-scores are higher. Similarly, low *ff*s (4% in some clinical laboratories)

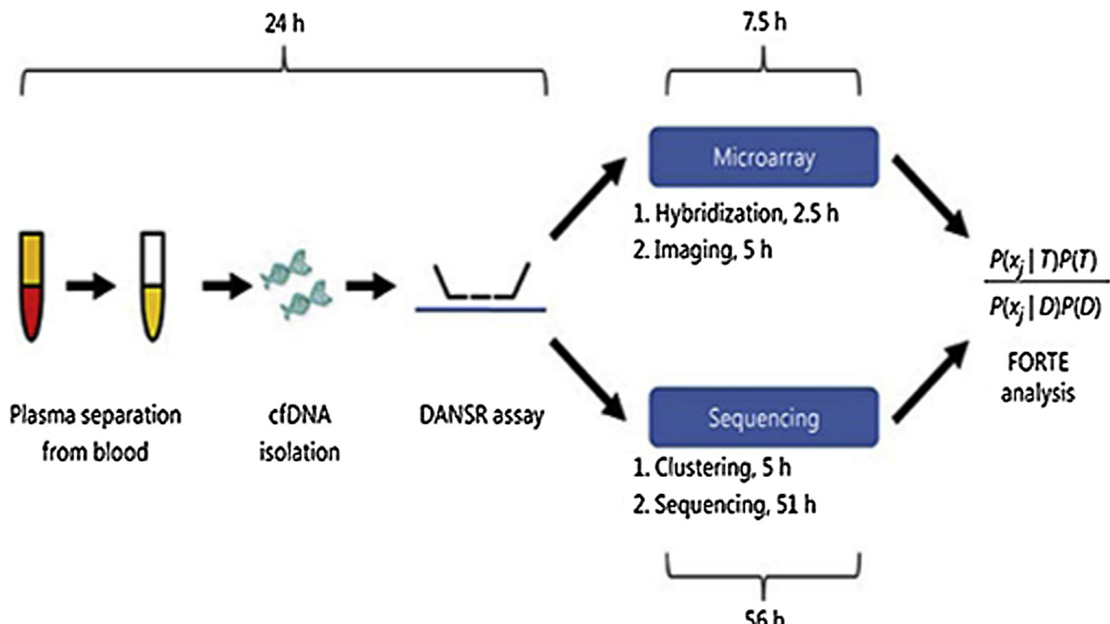


FIGURE 24.8 Microarray-based targeted methodology. *cfDNA*, cell-free DNA; *DANSR*, digital analysis of selected regions; *FORTE*, fetal-fraction optimized risk of trisomy evaluation.

are related to test failures and FN test results. Both maternal and fetal factors such as maternal weight and ethnicity, gestational age, fetal aneuploidy, twin pregnancy, and mosaicism affect *ff* and have to be taken into account during NIPT (Bianchi and Wilkins-Haug, 2014).

Maternal body mass index (BMI) is the most significant demographic variable that affects *ff* (Vora et al., 2012). The total circulating cfDNA is proportionately increased in pregnant women as a function of BMI. The fetal DNA is unaffected, but the increased maternal DNA results in an overall lower *ff*. As for gestational age, it has been shown that *ff* increases by 0.1% per week between 10 and 21 weeks of gestation ($P < .0001$), and after 21 weeks it increases by 1% per week ($P < .0001$). Together, maternal weight and fetal gestational age accounted for as much as 27% of the interindividual variation seen in *ff* (Wang et al., 2013). Relative to euploid males, fetuses with trisomy 21 have an increased *ff*, and fetuses with trisomy 18, 13, and monosomy X have a decreased *ff* (Rava et al., 2013).

NIPT is possible for twin pregnancies. If both twins have the same genotype, they can effectively be treated like a singleton gestation, and the *Z*-scores for the aneuploid chromosome should have a linear relationship with *ff*. The difficulty arises when twins are discordant for aneuploidy, because *ff* is not doubled. Although the overall *ff* is increased, the *ff* per fetus is reduced compared with singletons. In approximately 10−15% of cases, the *ff* would be lower than 4%, and therefore there would be a risk of a FN result (Srinivasan et al., 2013).

If there is placental or fetal mosaicism with a cell line that has autosomal aneuploidy, a higher overall *ff* will increase the chance that the partial DNA contribution from a mosaic abnormal line will be found, while low percentages of mosaicism (10%) cannot be detected (Canick et al., 2013).

24.6.2 Mosaicism

If there is placental or fetal mosaicism with a cell line that has autosomal aneuploidy, a higher overall *ff* will increase the chance that the partial DNA contribution from a mosaic abnormal line will be found (Canick et al., 2013). Similarly, a high percentage of abnormal cells will increase the chance that sequences from the aneuploidy chromosome will be detected. It has been shown that NIPT detected the aneuploidy in mosaic cases with at least 29% of cells positive for trisomies 18 and 21 (Bianchi et al., 2012). In contrast, low percentages of mosaicism (10%) or low effective *ff*s have led to FN results (Canick et al., 2013). It is important to take into consideration the different types of placental and fetal mosaicism in order to appraise the influence of mosaicism on NIPT results (Fig. 24.9) (Gardner et al., 2012). In cases where there is a discordance between the chromosomal constitution of the placenta and the fetus there is the risk of FP or FN NIPT results, since cffDNA is of placental origin only (Fig. 24.9B and G) (Koumbaris et al., 2016). The same applies in cases of placental or fetal mosaicism (Fig. 24.9C−F). Confined placental mosaicism (CPM) (Fig. 24.9B) is found in approximately 1−2% of chorionic villi samples (Hahnemann et al., 1997; Gardner et al., 2012) and is the most frequent reason for FP NIPT results, especially involving the sex chromosomes (Brady et al., 2016; Koumbaris et al., 2016).

24.6.3 Chromosome Guanine Cytosine Base Content

Fan and Quake (2010) showed that sequencing results were biased in chromosomes with increased or decreased GC base content levels. They described a computational method to remove the GC bias. This eliminated the variance in the distribution of sequence tag counts and increased statistical confidence in the detection of aneuploidy. They also showed that with increased sequence tags there was an increased precision of calling a sample aneuploidy. After GC correction, MPS counting statistics follow a normal distribution. Each testing company has different proprietary algorithms to determine cutoffs between normal and abnormal. The classification cutoff will determine the expected FPRs.

Chromosomes that have GC content in the mid-range (such as 21, 18, and X) perform best with MPS (Chiu et al., 2008). Chromosome 13 has one of the lowest GC contents. For this reason, specific approaches have been used to normalize the GC ratios for chromosome 13, which have resulted in an increased sensitivity of detection of trisomy 13 (Fan et al., 2010; Chen et al., 2011).

24.6.4 Chromosomal Value

An additional factor that affects sensitivity is the chromosomal value (CV) in the sequencing results for a specific chromosome. Normalized chromosome value (NCV) is a mathematical calculation for comparing each chromosome tested in cfDNA (Sehnert et al., 2011). NCV calculation removes variation within and between sequencing runs to optimize test precision. An excess or deficiency in the number of counts, expressed as a *Z*-score or NCV, implies aneuploidy for that chromosome. A smaller CV results in larger *Z*-scores, making it easier to separate aneuploid from euploid samples. Chromosome 18 has approximately half of the CV observed for chromosomes 13, 21, and X (Rava et al., 2013). The clinical implication is that trisomy 18 can be detected at lower *ff*s. It is therefore easier to detect low-grade mosaicism, which may be one reason why there are more discordant results between NIPT and karyotype with trisomy 18 (Mennuti et al., 2013).

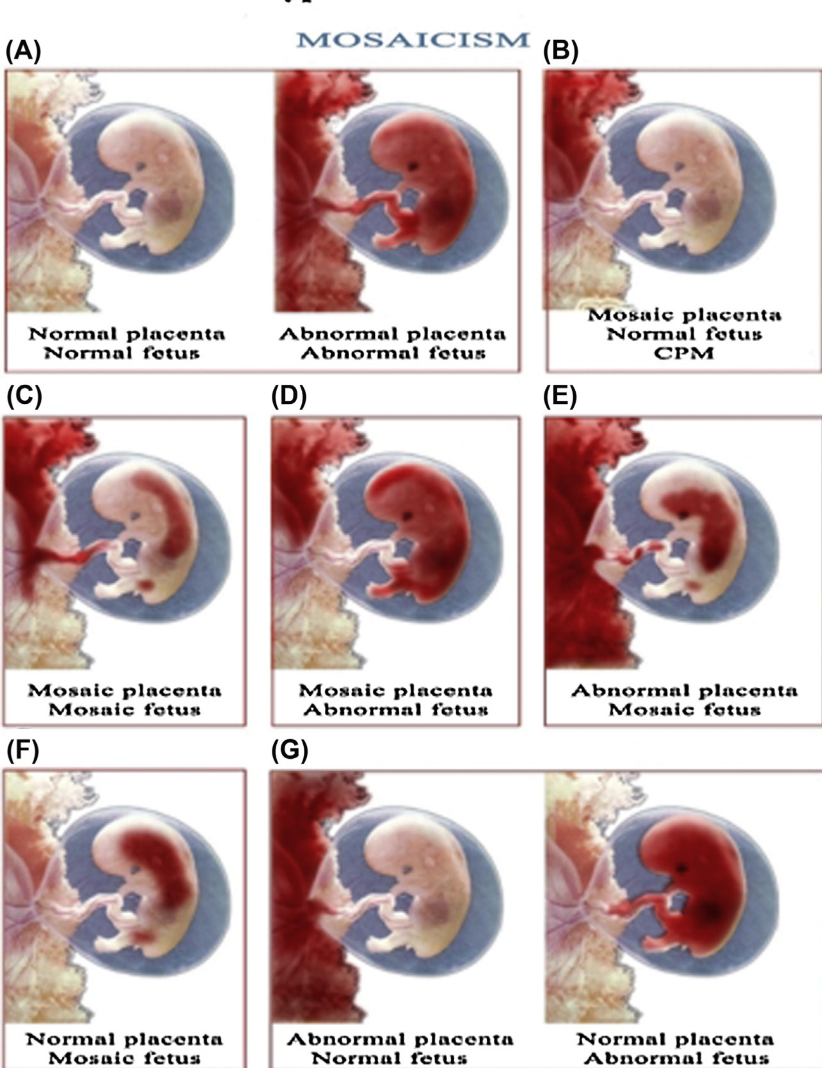


FIGURE 24.9 Types of mosaicism: (A) total agreement between the chromosomal constitution of the placenta and the fetus; (B) confined placental mosaicism; (C) total agreement for mosaicism between the placenta and the fetus; (D, E) partial agreement for mosaicism between the placenta and the fetus; (F) confined fetal mosaicism; (G) total discordance between the chromosomal constitution of the placenta and the fetus. *CPM*, confined placental mosaicism.

24.6.5 Maternal Copy Number Variations

Maternal CNVs may lead to FP and FN NIPT results because NIPT relies upon the analysis of cfDNA derived from the maternal plasma, and the majority of cfDNA is maternal in origin. Therefore counting the statistics of conventional *Z*-score chromosome-wide analysis methods can be affected by maternal CNVs. The development and clinical application of analysis pipelines, which allow for differentiation between localized and chromosome-wide events, including maternal CNVs, may avoid such errors (Bayindir et al., 2015). One study found that maternal CNVs were a major factor contributing to FP and FN NIPT results as well as fetal/placental mosaicism (Zhang et al., 2015). These findings demonstrate the importance of differentiating between whole chromosome and sub-chromosomal events as well as those that are of maternal origin.

24.7 NONINVASIVE PRENATAL TESTING IN CLINICAL TRIALS

In a metaanalysis of the performance of cffDNA testing when screening for aneuploidies, a total of 37 relevant studies were identified (Gil et al., 2015). These studies reported cffDNA results in relation to fetal karyotype from

TABLE 24.1 Data From a Metaanalysis Study on the Performance of Cell-Free Fetal DNA Testing in Screening for Aneuploidies in a Total of 37 Relevant Studies in Singleton Pregnancies

	Detection Rate (%)	False Positive (%)	False Negative (%)
Trisomy 21	99.2	0.09	0.36
Trisomy 18	96.3	0.13	1.9
Trisomy 13	91	0.13	6.67
Monosomy X	90.3	0.23	
XXX, XXY	93	0.14	

Reproduced with permission from Gil, M.M., Quezada, M.S., Revello, R., Akolekar, R., Nicolaides, K.H., 2015. Analysis of cell-free DNA in maternal blood in screening for fetal aneuploidies: updated meta-analysis. Ultrasound Obstet. Gynecol. 45 (3), 249–266.

invasive testing or clinical outcome. Weighted pooled DRs and FPRs for singleton pregnancies are shown in Table 24.1.

24.7.1 Performance of Screening for Aneuploidies

As can be seen in Table 24.1, in singleton pregnancies, cffDNA analysis of maternal blood can detect more than 99% of cases of fetal trisomy 21 with an FPR of less than 0.1%. In the metaanalysis by Gil et al. (2015), the combined total number of affected ($n = 1051$) and unaffected ($n = 21{,}608$) pregnancies was large, and the heterogeneity between studies was low. Although most studies were in high-risk pregnancies, there were five studies with a combined total of 57 affected and 8685 unaffected pregnancies in general populations with a DR of 100% and an FPR of 0.08%.

The performance of cffDNA analysis of maternal blood in the identification of singleton pregnancies with fetal trisomy 18 or 13, with respective DRs of about 96% and 91% and a combined FPR of 0.26%, is worse than the performance of screening for trisomy 21. The objective of trying to identify all three trisomies, rather than trisomy 21 alone, is achieved at the expense of a fourfold increase in the FPR, from 0.09% to 0.35%. Furthermore, the number of affected cases examined, 389 for trisomy 18 and 139 for trisomy 13, was considerably smaller than that for trisomy 21, and the heterogeneity in DR and FPR between studies was much higher for trisomy 13 than for the other two trisomies.

A small number of studies, with a combined total of 177 singleton pregnancies with fetal monosomy X and 56 with other sex chromosome aneuploidies, reported that cffDNA analysis of maternal blood detected about 90% of the former and 93% of the latter, with a combined FPR of 0.37%. Notably, in some studies the rate of laboratory failure to provide a result was considerably higher for sex chromosome aneuploidies than it was for the trisomies.

In another metaanalysis study by Taylor-Phillips et al. (2016) after a systematic review of 2012 articles, the test accuracy of NIPT pooled sensitivity was 99.3% for trisomy 21, 97.4% for trisomy 18, and 97.4% for trisomy 13, with a pooled specificity of 99.9% (99.9–100%) for all three trisomies (Table 24.2).

24.7.2 Screening for Aneuploidies in Twin Pregnancies

While screening by cffDNA testing is feasible in twin pregnancies, the performance of the screening may be worse than it is in singletons. The DR for trisomy 21 is 93.7%, and the FPR is 0.23% (Gil et al., 2015). In the metaanalysis by Taylor-Phillips et al. (2016) the sensitivity was also lower in twin than singleton pregnancies, reduced by 9% for Down, 28% for Edwards, and 22% for Patau syndrome.

In twins, cffDNA testing is more complex, because the two fetuses could be either monozygotic, and therefore

TABLE 24.2 Data From a Metaanalysis Study on the Performance of Cell-Free Fetal DNA Testing in Screening for Aneuploidies in a Total of 41, 37, and 30 Studies in the Review for Trisomy 21, Trisomy 18, and Trisomy 13, Respectively, in Singleton Pregnancies

	Sensitivity (%)	Specificity (%)
Trisomy 21	99.3	99.9–100
Trisomy 18	97.4	99.9–100
Trisomy 13	97.4	99.9–100

Reproduced with permission from Taylor-Phillips, S., Freeman, K., Geppert, J., Agbebiyi, A., Uthman, O.A., Madan, J., Clarke, A., Quenby, S., Clarke, A., January 18, 2016. Accuracy of non-invasive prenatal testing using cell-free DNA for detection of Down, Edwards and Patau syndromes: a systematic review and meta-analysis. BMJ Open 6 (1).

genetically identical, or dizygotic, in which case only one fetus is likely to have any aneuploidy identified. There is evidence that, in dizygotic twins, each fetus can contribute different amounts of cffDNA into the maternal circulation, and the difference can be nearly twofold. It is therefore possible, in a dizygotic twin pregnancy discordant for aneuploidy, for the *ff* of the affected fetus to be below the threshold (4%) for successful cffDNA testing. This could lead to an erroneous result of low risk for aneuploidy, with a high contribution from the disomic co-twin resulting in a satisfactory total *ff*. To avoid this potential error, it was proposed that for cffDNA testing in twin pregnancies, the lower *ff* of the two fetuses, rather than the total *ff*, should be estimated in the assessment of risk for aneuploidies. However, an inevitable consequence of such a policy is that the no-result rate in twins is higher than in singleton pregnancies.

24.7.3 False Positive and False Negative Results

The FPRs for NIPT are on the order of 0.1–0.2%. It remains unknown, however, what portion of the discordancy between NIPT and fetal results is attributable to the demarcation between euploidy and aneuploidy established by the statistical algorithms and how much is due to biologic explanations uncovered by the greater sensitivity of NGS. With regard to FN cases, the few reported cases suggest that the underlying explanation is either a sample mix-up, low *ff*, or mosaicism (Fig. 24.9F and G) (Canick et al., 2013; Futch et al., 2013).

CPM, as already mentioned, is present in 1–2% of first-trimester placentas (Fig. 24.9B) (Hahnemann et al., 1997; Gardner et al., 2012). Several instances of CPM detected at CVS have been reported in association with discordant positive NIPT results (Mennuti et al., 2013; Hall et al., 2013).

The frequency and significance of CPM are relevant to NIPT because cffDNA is believed to originate primarily from nonviable trophoblasts. CPM could result in either FP or FN NIPT results. It has been suggested that viable pregnancies with Edwards (trisomy 18) or Patau (trisomy 13) syndrome may have a substantial euploid cell line in the trophoblasts (Kalousek et al., 1989; Schuring-Blom et al., 2001), and therefore NIPT could fail to detect some viable trisomic cases in which there is a substantial population of abnormal cells in the fetus but not in the placenta (Fig. 24.9C and D). This would lead to an apparent FN result by NIPT. Conversely, the trophoblasts may be mostly abnormal and the fetal cells normal, leading to an apparent FP result (Fig. 24.9B and F). Moreover, if chromosomally abnormal trophoblasts were more likely to undergo apoptosis, NIPT could preferentially identify cases in which the proportion of abnormal cells in the placenta is low. Therefore as NIPT tests are expanded to include additional chromosomes or chromosomal regions, the issue of erroneous results due to CPM is likely to become increasingly important.

For those cases that do show true fetal mosaicism (Fig. 24.9C and E), the phenotype can be highly variable but will generally be less severe than that present with the nonmosaic abnormal karyotype (Benn, 2010). Although the proportion of abnormal cells in the amniotic fluid or chorionic villi can be a poor guide to phenotype in individual cases, in general, a high proportion of abnormal cells is associated with an adverse pregnancy outcome (Wallerstein et al., 2000). NIPT tests that only present a result categorically as either "positive" or "negative" cannot provide any information about the proportion of abnormal cells or clinical severity. Furthermore, since some of the clinical trials have specifically excluded mosaic cases and optimized their *Z*-score cutoff on the basis of nonmosaic cases, testing is presumably biased toward not detecting mosaic cases.

In a singleton pregnancy that had a vanishing twin, it is theoretically possible that apoptosis of cells from the fetoplacental remains of the nonviable fetus could interfere with the cffDNA result; there is a relatively high chance that the fetal loss would have been associated with aneuploidy, and this could lead to an FP result (Futch et al., 2013).

Maternal conditions should also be considered as possible biologic explanations of NIPT-fetal discordant results. These include constitutional maternal aneuploidies, most likely involving sex chromosomes, and conditions with intrinsic genomic alterations, such as solid tumors (Osborne et al., 2013).

The initial report of a maternal sex chromosome aneuploidy (SCA) occurred in a completely normal 25-year-old pregnant woman whose NIPT results were consistent with triple X (Yao et al., 2012). Amniocentesis revealed a fetal 46,XX karyotype; her infant was normal. The maternal blood karyotype showed full 47,XXX. Her presentation emphasizes the wide range of phenotypic variability with the SCAs. SCAs, especially mosaic cases, are likely to be underappreciated clinically. The possibility of their identification through NIPT requires discussion during counseling.

24.8 NONINVASIVE PRENATAL TESTING IN THE CLINICAL SETTING

There is clear evidence that in singleton pregnancies the performance of screening for trisomy 21 by cffDNA testing is superior to that of all other methods combining maternal age, first- or second-trimester ultrasound findings, and first- or second-trimester serum biochemical

analysis. Additionally, the test can be carried out at 10–11 weeks' gestation, with the advantage of providing early reassurance for the majority of parents that their fetus is unlikely to be trisomic and, for the few with an affected fetus, the parents have the option of an earlier and safer termination of pregnancy (Gil et al., 2013; Quezada et al., 2015).

NIPT could be implemented into this testing pipeline in one of three ways: as a replacement for serum screening, as an intermediate step between screening and invasive procedures, or as a replacement for invasive testing. A sensitivity analysis in which the factors contributing to the overall costs of four different cffDNA screening policies has been published by Cuckle et al. (2013). However, according to Gil et al. (2015), there are essentially two options in the clinical implementation of cffDNA analysis of maternal blood in screening for trisomy 21: routine screening of the whole population and contingent screening based on the results of first-line screening by another method, preferably the first-trimester combined test (Gil et al., 2015). The two major limitations of cffDNA testing as a potential method for universal screening are the high cost of the test and the rate of failure to provide a result. According to Gil et al. (2014, 2015), both of these problems can be overcome by the use of cffDNA testing, depending on the results of the first-trimester combined test contingent screening which would lead to a very high DR and a very low invasive testing rate at a considerably lower cost than compared with carrying out cffDNA testing as a first-line method of screening. In cases of failed cffDNA testing, pregnant women can rely on the results of the combined test in deciding in favor or against invasive testing. This strategy would also retain the advantages of first-trimester testing by ultrasound and biochemistry, including accurate pregnancy dating, early detection of many major fetal defects, and prediction, with the potential of prevention, of a wide range of pregnancy complications, including preeclampsia and preterm birth (Nikolaides, 2011; Nikolaides et al., 2012).

As in traditional prenatal testing, lethal trisomies 18 and 13 have been included in NIPT, although studies on cffDNA testing suggest that the performance of screening for trisomies 18 and 13 may be worse than that of the combined test (Gil et al., 2015). Large studies utilizing the first-trimester combined test have reported that the use of risk algorithms for each of the three trisomies results in DRs of about 90% for trisomy 21% and 95% for trisomies 18 and 13, with an increase in FPR of only 0.1% above the FPR of about 4% in screening for trisomy 21 alone (Wright et al., 2014). Although the reported DR of the two tests is similar, it is likely that the true DR of the cffDNA test will be lower if the cases in which the test fails to give a result are included. Furthermore, the differential increase in FPR by including these trisomies in a screening strategy aimed primarily at detecting trisomy 21 is considerably higher with cffDNA testing than with the combined test.

Since both trisomy 13 and trisomy 18 pregnancies are highly likely to have abnormalities detectable by ultrasound investigation and will spontaneously abort between 12 weeks and term, the main concern remains for FP and FN trisomy 21 results. Taking into consideration the incidence of trisomy 21 in the general population and the incidence of true fetal mosaicism type V, the number of FN trisomy 21 cases is estimated to be approximately 1 in 100,000 NIPTs (Koumbaris et al., 2016). Also, assuming that at least 70% CPM is needed in order to produce an FP trisomy 21 result, the FP trisomy 21 rate would be approximately 1 in 13,000. Although these figures are very low, it is most important to be aware of the genetic physiology of the placenta and the limitations it imposes on NIPT when contemplating its integration into safe clinical prenatal care.

As for NIPT of SCA, it is debatable whether it should be offered, because it generally results in a phenotype which is mild, without physical or intellectual disability. The only exception is the lethal type of monosomy X (Turner syndrome), which presents with a very large nuchal translucency (NT) during the first trimester or cystic hygroma/hydrops during the second trimester; in such cases the investigation of choice would be invasive testing for fetal karyotype evaluation, including subchromosomal analysis with microarray, rather than cffDNA testing for assessment of risk for 45,X. Another reason for not offering SCA NIPT is that the FPR is high, because CPM is frequent in this type of aneuploidy (Gardner et al., 2012).

24.9 COUNSELING AND ETHICAL ISSUES

For any NIPT, caution is needed when it comes to genetic counseling in order to avoid misunderstandings concerning diagnosis. The counseling of women who are considering NIPT is challenging. The range of cytogenetic abnormalities currently detectable through cffDNA testing is smaller than that detectable by conventional karyotyping and substantially less than that achievable through microarray testing.

In the absence of public health policies that define testing strategies, each high-risk woman who previously would have made a decision about amniocentesis or CVS and karyotyping will now need to choose between no testing, NIPT, invasive testing with conventional cytogenetics, or invasive testing with microarray testing. These women will need to be counseled carefully regarding the complex benefits and hazards associated with each option.

As long as NIPT is used to only test for Down syndrome and trisomies 18 and 13, introducing this new test will result in detecting fewer chromosomal abnormalities than the traditional approach, and this must be carefully explained to the woman (Petersen et al., 2014). Some of

these additional findings are serious conditions, not all of which will come to light with the second-trimester fetal anomaly scan. Although the risk of missing a clinically relevant abnormality is actually quite low (with an estimated range between 1:1600 and 1:4000) (Verweij et al., 2014), it has been suggested that the benefits of NIPT may not outweigh the loss of these extra findings at the stage of invasive follow up (Benn et al., 2011). The availability of professional guidelines can greatly facilitate counseling by clarifying which processes are scientifically valid and clinically acceptable standards of care (Benn et al., 2013).

The available NIPT for aneuploidy is so far only provided by private companies. It is important to develop an understanding of the potential harms and benefits of intellectual property, which is grounded in stakeholders' perspectives and the regulatory and legal environment. Understanding intellectual property and commercialization-related barriers to clinical adoption and patient access early in the translation pathway can guide policies to ensure the efficient implementation of NIPT and appropriate access by all patient populations (Agarwal et al., 2013).

There is an ongoing debate on ethical and policy issues related to NIPT, and both the European and American Human Genetics Societies have contributed to setting future guidelines for NIPT (Dondorp et al., 2015).

24.10 FUTURE APPLICATIONS OF NONINVASIVE PRENATAL TESTING

Commercially available NIPT methodologies present high-accuracy classification results. However, NIPT-positive results should be confirmed by invasive amniocentesis, as CPM and other rare molecular events may provide false results. This means that a probability is given for each condition investigated, and depending on whether the pregnancy is assessed as being high risk or not, the couples are counseled to proceed with confirmatory invasive diagnostic testing, usually fetal karyotyping after an amniocentesis. FP results lead to unneeded invasive procedures, posing an undesirable risk of pregnancy loss, while FN results may lead to the birth of an abnormal child. There is an argument that FN NIPT results for trisomy 18 or 13 are unlikely to result in the birth of an abnormal child, because both syndromes are most likely to present with serious ultrasound findings during pregnancy. Conversely, cases with trisomy 21 (Down syndrome) may not have any indications throughout the pregnancy, and consequently, NIPT FN trisomy 21 fetuses are more likely to be born (Smith et al., 2014). However, as it has already been mentioned, the number of FN trisomy 21 cases is estimated to be approximately 1 in 100,000 NIPTs (Koumbaris et al., 2016), making the test near diagnostic when all necessary methodology requirements are met and appropriate biological parameters are taken into account.

If NIPT is treated as a replacement for current biochemical screening tests, the resulting risk could be combined with that derived from ultrasound markers such as NT measurement and others. The combined NIPT—ultrasound risk for fetal aneuploidy may provide a safer screening strategy compared to that offered to most couples today (Wright et al., 2015). The future aim is to eventually avoid invasive procedures and evolve NIPT (testing) into NIPD (diagnosis).

Methodologies based on NGS are extremely powerful. Besides detecting whole chromosome aneuploidy, they have the potential to detect smaller chromosomal imbalances, allowing for microdeletion/microduplication syndrome NIPT (Hahn et al., 2011). However, although NIPT is already commercially available for the detection of a certain number of microdeletion/microduplication syndromes, further validation studies are needed (Benn and Cuckle, 2014). Taking into account the vast amount of data that NGS is capable of producing, it could potentially be combined with other methodologies to generate noninvasive fetal WGS (Lo, 2013). As impressive as this may seem at present, it is quite possible that this will materialize in the near future.

24.11 CONCLUSIONS

The pace of NIPT is unprecedented in clinical laboratory medicine. The past years have seen rapid advances in fetal genomic analyses from maternal plasma, and this will hopefully translate equally rapidly to advances in patient care. The field began to develop with the discovery of cffDNA in the maternal plasma in 1997 (Lo et al., 1997). This was followed by the discovery of placental DNA methylation markers on chromosome 21 for the development of an epigenetic test for fetal trisomy 21 in 2006 (Chan et al., 2006). In 2007, digital PCR for NIPT for trisomy 21 was described (Fan and Quake, 2007), while the development of fetal aneuploidy testing by massively parallel sequencing (MPS) in 2008 (Chiu et al., 2008) paved the way for the implementation of NIPT for fetal aneuploidy as a clinical service in 2011 (Palomaki et al., 2011, 2012; Bianchi et al., 2012; Norton et al., 2012). All studies to date indicate that NIPT produces lower FPs and higher positive predictive values than serum screening. Since a major objective in the field of prenatal testing is the reduction of the number of unnecessary invasive procedures, NIPT can significantly reduce procedure-related losses while maintaining high DRs. It provides clinicians and prospective parents with a powerful tool to help them make informed decisions regarding the need for an invasive procedure, without posing any risk to the pregnancy.

The clinical impact of NIPT has been significant, as indicated by its quick adoption in prenatal care. The extent to which it can be applied as a universal screening tool for trisomy 21, 18, and 13 depends mainly on assay accuracy, a low number of nonreportable tests, and cost. One survey indicated that clinicians believe that NIPT will someday replace conventional screening procedures (Haymon et al., 2014). With further advances in technology and reductions in costs, it is possible that noninvasive prenatal genome-wide analysis will play an increasingly important role in the future practice of prenatal medicine.

CONFLICT OF INTEREST

Philippos C. Patsalis declares a conflict of interest as he owns shares in NIPD Genetics Ltd.

REFERENCES

Agarwal, A., Sayres, L.C., Cho, M.K., Cook-Deegan, R., Chandrasekharan, S., 2013. Commercial landscape of noninvasive prenatal testing in the United States. Prenat. Diagn. 33, 521−531. Review.

Alberry, M., Maddocks, D., Jones, M., Abdel-Hadi, M., Abdel-Fattah, S., Avent, N., Soothill, B.W., 2007. Free fetal DNA in maternal plasma in anembryonic pregnancies: confirmation that the origin is the trophoblast. Prenat. Diagn. 27, 415−418.

Alfirevic, Z., Sundberg, K., Brigham, S., 2003. Amniocentesis and chorionic villus sampling for prenatal diagnosis. Cochrane Database Syst. Rev. 3, CD003252.

Allyse, M., Minea, M.A., Berson, E., Sridhar, S., Rote, M., Hung, A., Chandrasekharan, S., 2015. Non-invasive prenatal testing: a review of international implementation and challenges. Int. J. Womens Health 7, 113−126.

Ashoor, G., Syngelaki, A., Wang, E., Struble, C., Oliphant, A., Song, K., Nicolaides, K.H., 2013. Trisomy 13 detection in the first trimester of pregnancy using a chromosome-selective cell-free DNA analysis method. Ultrasound Obstet. Gynecol. 41, 21−25.

Bayindir, B., Dehaspe, L., Brison, N., et al., 2015. Non-invasive prenatal testing (NIPT) using a novel analysis pipeline to screen for all autosomal fetal aneuploidies improves pregnancy management. Eur. J. Hum. Genet. 23, 1286−1293.

Benn, P., Borrell, A., Crossley, J., Cuckle, H., Dugoff, L., Gross, S., Johnson, J.A., Maymon, R., Odibo, A., Schielen, P., Spencer, K., Wright, D., Yaron, Y., 2011. Aneuploidy screening: a position statement from a committee on behalf of the board of the international society for prenatal diagnosis. Prenat. Diagn. 31, 519−522.

Benn, P., Cuckle, H., 2014. Theoretical performance of non-invasive prenatal testing for chromosome imbalances using counting of cell-free DNA fragments in maternal plasma. Prenat. Diagn. 34, 778−783.

Benn, P.A., Cuckle, H., Pergament, E., 2013. Non-invasive prenatal testing for aneuploidy: current status and future prospects. Ultrasound Obstet. Gynecol. 42, 15−33.

Benn, P.A., 2010. Prenatal diagnosis of chromosome abnormalities through amniocentesis. In: Milunsky, A., Milunsky, J.M. (Eds.), Genetic Disorders and the Fetus: Diagnosis, Prevention and Treatment, sixth ed. Wiley-Blackwell, Chichester, UK, pp. 771−818.

Bianchi, D.W., Wilkins-Haug, L., 2014. Integration of noninvasive DNA testing for aneuploidy into prenatal care: what has happened since the rubber met the road? Clin. Chem. 60 (1), 78−87. Review.

Bianchi, D.W., Simpson, J.L., Jackson, L.G., Elias, S., Holzgreve, W., Evans, M.I., Dukes, K.A., Sullivan, L.M., Klinger, K.W., Bischoff, F.Z., Hahn, S., Johnson, K.L., Lewis, D., Wapner, R.J., de la Cruz, F., 2002. Fetal gender and aneuploidy detection using fetal cells in maternal blood: analysis of NIFTY I data. National Institute of Child Health and Development Fetal Cell Isolation Study. Prenat. Diagn. 22, 609−615.

Bianchi, D.W., Platt, L.D., Goldberg, J.D., Abuhamad, A.Z., Sehnert, A.J., Rava, R.P., MatErnal BLood IS Source to Accurately Diagnose Fetal Aneuploidy (MELISSA) Study Group, 2012. Genome-wide fetal aneuploidy detection by maternal plasma DNA sequencing. Obstet. Gynecol. 119, 890−901.

Bianchi, D.W., 2004. Circulating fetal DNA: its origin and diagnostic potential − a review. Placenta 25, S93−S101.

Brady, P., Brison, N., Van Den Bogaert, K., de Ravel, T., Reeters, H., Van Esch, H., Devriendt, K., Legius, E., Vermeesch, J.R., 2016. Clinical implementation of NIPT − technical and biological challeges. Clin. Genet. 89, 523−530.

Caine, A., Maltby, A.E., Parkin, C.A., Waters, J.J., Crolla, J.A., UK Association of Clinical Cytogeneticists (ACC), 2005. Prenatal detection of Down's syndrome by rapid aneuploidy testing for chromosomes 13, 18, and 21 by FISH or PCR without a full karyotype: a cytogenetic risk assessment. Lancet 366 (9480), 123−128.

Canick, J.A., Palomaki, G.E., Kloza, E.M., Lambert-Messerlian, G.M., Haddow, J.E., 2013. The impact of maternal plasma DNA fetal fraction on next generation sequencing tests for common fetal aneuploidies. Prenat. Diagn. 33, 667−674.

Chan, K.C., Ding, C., Gerovassili, A., Yeung, S.W., Chiu, R.W., Leung, T.N., Lau, T.K., Chim, S.S., Chung, G.T., Nicolaides, K.H., Lo, Y.M., 2006. Hypermethylated RASSF1A in maternal plasma: a universal fetal DNA marker that improves the reliability of noninvasive prenatal diagnosis. Clin. Chem. 52, 2211−2218.

Chen, E.Z., Chiu, R.W., Sun, H., Akolekar, R., Chan, K.C., Leung, T.Y., Jiang, P., Zheng, Y.W., Lun, F.M., Chan, L.Y., Jin, Y., Go, A.T., Lau, E.T., To, W.W., Leung, W.C., Tang, R.Y., Au-Yeung, S.K., Lam, H., Kung, Y.Y., Zhang, X., van Vugt, J.M., Minekawa, R., Tang, M.H., Wang, J., Oudejans, C.B., Lau, T.K., Nicolaides, K.H., Lo, Y.M., 2011. Noninvasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma DNA sequencing. PLoS One 6, e21791.

Chim, S.S., Jin, S., Lee, T.Y., Lun, F.M., Lee, W.S., Chan, L.Y., Jin, Y., Yang, N., Tong, Y.K., Leung, T.Y., Lau, T.K., Ding, C., Chiu, R.W., Lo, Y.M., 2008. Systematic search for placental DNA methylation markers on chromosome 21: toward a maternal plasma based epigenetic test for fetal trisomy 21. Clin. Chem. 54, 500−511.

Chiu, R.W., Chan, K.C., Gao, Y., Lau, V.Y., Zheng, W., Leung, T.Y., Foo, C.H., Xie, B., Tsui, N.B., Lu, F.M., Zee, B.C., Lau, T.K., Cantor, C.R., Lo, Y.M., 2008. Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. Proc. Natl. Acad. Sci. U.S.A. 105, 20458−20463.

Chiu, R.W.K., Akolekar, R., Zheng, Y.W., Leung, T.Y., Sun, H., Chan, K.C., Lun, F.M., Go, A.T., Lau, E.T., To, W.W., Leung, W.C., Tang, R.Y., Au-Yeung, S.K., Lam, H., Kung, Y.Y., Zhang, X., van Vugt, J.M., Minekawa, R., Tang, M.H., Wang, J., Oudejans, C.B.,

Lau, T.K., Nicolaides, K.H., Lo, Y.M., 2011. Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study. BMJ 342, c7401.

Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V., Abdel-Hamid, H., Bader, P., McCracken, E., Niyazov, D., Leppig, K., Thiese, H., Hummel, M., Alexander, N., Gorsk, J., Kussmann, J., Shashi, V., Johnson, K., Rehder, C., Ballif, B.C., Shaffer, L.G., Eichler, E.E., 2011. A copy number variation morbidity map of developmental delay. Nat. Genet. 43, 838–846.

Cuckle, H., Benn, P., Pergament, E., 2013. Maternal cfDNA screening for Down's syndrome—a cost sensitivity analysis. Prenat. Diagn. 33 (7), 636–642.

Dan, S., Wang, W., Ren, J., Li, Y., Hu, H., Xu, Z., Lau, T.K., Xie, J., Zhao, W., Huang, H., Sun, L., Zhang, X., Wang, W., Liao, S., Qiang, R., Cao, J., Zhang, Q., Zhou, Y., Zhu, H., Zhong, M., Guo, Y., Lin, L., Gao, Z., Yao, H., Zhang, H., Zhao, L., Jiang, F., Chen, F., Jiang, H., Li, S., Li, Y., Wang, J., Wang, J., Duan, T., Su, Y., Zhang, X., 2012. Clinical application of massively parallel sequencing-based prenatal noninvasive fetal trisomy test for trisomies 21 and 18 in 11,105 pregnancies with mixed risk factors. Prenat. Diagn. 32, 1225–1232.

de Wit, M.C., Srebniak, M.I., Govaert, L.C., Van Opstal, D., Galjaard, R.J., Go, A.T., 2014. Additional value of prenatal genomic array testing in fetuses with isolated structural ultrasound abnormalities and a normal karyotype: a systematic review of the literature. Ultrasound Obstet. Gynecol. 43, 139–146.

Dhallan, R., Guo, X., Emch, S., Damewo, M., Bayliss, P., Cronin, M., Barry, J., Betz, J., Franz, K., Gold, K., Vallecillo, B., Varney, J., 2007. A non-invasive test for prenatal diagnosis based on fetal DNA present in maternal blood: a preliminary study. Lancet 369, 474–481.

Dondorp, W., de Wert, G., Bombard, Y., Bianchi, D.W., Bergmann, C., Borry, P., Chitty, L.S., Fellmann, F., Forzano, F., Hall, A., Henneman, L., Howard, H.C., Lucassen, A., Ormond, K., Peterlin, B., Radojkovic, D., Rogowski, W., Soller, M., Tibben, A., Tranebjærg, L., van El, C.G., Cornel, M.C., 2015. Non-invasive prenatal testing for aneuploidy and beyond: challenges of responsible innovation in prenatal screening. Summary and recommendations. Eur. J. Hum. Genet. 1–3.

Evangelidou, P., Patsalis, P.C., Ioannides, M., Koumbaris, G., Kallikas, I., Velissariou, V., Sismani, C., 2013. Implementation of high resolution whole genome array CGH in the prenatal clinical setting. Advantages, challenges and review of the literature. Biomed. Res. Int. 2013, 346762.

Faas, B.H., de Ligt, J., Janssen, I., Eggink, A.J., de Wijnberger, L.D., van Vugt, J.M., Vissers, L., Geurts van Kessel, A., 2012. Non-invasive prenatal diagnosis of fetal aneuploidies using massively parallel sequencing-by-ligation and evidence that cell-free fetal DNA in the maternal plasma originates from cytotrophoblastic cells. Expert Opin. Biol. Ther. 12 (Suppl. 1), S19–26.

Fan, H.C., Quake, S.R., 2007. Detection of aneuploidy with digital polymerase chain reaction. Anal. Chem. 79, 7576–7579.

Fan, H.C., Quake, S.R., 2010. Sensitivity of noninvasive prenatal detection of fetal aneuploidy from maternal plasma using shotgun sequencing is limited only by counting statistics. PLoS One 5, e10439.

Fan, H.C., Blumenfeld, Y.J., Chitkara, U., Hudgins, L., Quake, S.R., 2008. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. Proc. Natl. Acad. Sci. U.S.A. 105, 16266–16271.

Futch, T., Spinosa, J., Bhatt, S., de Feo, E., Rava, R.P., Sehnert, A.J., 2013. Initial clinical laboratory experience in noninvasive prenatal testing for fetal aneuploidy from maternal plasma DNA samples. Prenat. Diagn. 33, 569–574.

Gardner, R.J.M., Sutherland, G.R., Shaffer, L.G., 2012. Chromosome Abnormalities and Genetic Counseling. Oxford Monographs on Medical Genetics No. 61. Oxford University Press.

Gerdes, T., Kirchhoff, M., Lind, A.M., Vestergaard Larsen, G., Kjaergaard, S., 2008. Multiplex ligation-dependent probe amplification (MLPA) in prenatal diagnosis-experience of a large series of rapid testing for aneuploidy of chromosomes 13, 18, 21, X, and Y. Prenat. Diagn. 28, 1119–1125.

Gil, M.M., Quezada, M.S., Bregant, B., Ferraro, M., Nicolaides, K.H., 2013. Implementation of maternal blood cell-free DNA testing in early screening for aneuploidies. Ultrasound Obstet. Gynecol. 42, 34–40.

Gil, M.M., Akolekar, R., Quezada, M.S., Bregant, B., Nicolaides, K.H., 2014. Analysis of cell-free DNA in maternal blood in screening for aneuploidies: meta-analysis. Fetal Diagn. Ther. 35, 156–173.

Gil, M.M., Quezada, M.S., Revello, R., Akolekar, R., Nicolaides, K.H., 2015. Analysis of cell-free DNA in maternal blood in screening for fetal aneuploidies: updated meta-analysis. Ultrasound Obstet. Gynecol. 45 (3), 249–266.

Hahn, S., Huppertz, B., Holzgreve, W., 2005. Fetal cells and cell free fetal nucleic acids in maternal blood: new tools to study abnormal placentation? Placenta 26, 515–526.

Hahn, S., Lapaire, O., Tercanli, S., Kolla, V., Hösli, I., 2011. Determination of fetal chromosome aberrations from fetal DNA in maternal blood: has the challenge finally been met? Expert Rev. Mol. Med. 13, e16. http://dx.doi.org/10.1017/S1462399411001852. PMID:21542948.

Hahnemann, J.M., Vejerslev, L.O., 1997. Accuracy of cytogenetic findings on chorionic villus sampling (CVS) – diagnostic consequences of CVS mosaicism and non-mosaic discrepancy in centres contributing to EUCROMIC 1986–1992. Prenat. Diagn. 17, 801–820.

Hall, A.L., Drendel, H.M., Verbrugge, J.L., Reese, A.M., Schumacher, K.L., Griffith, C.B., Weaver, D.D., Abernathy, M.P., Litton, C.G., Vance, G.H., 2013. Positive cell free fetal DNA testing for trisomy 13 reveals confined placental mosaicism. Genet. Med. 15, 729–732.

Haymon, L., Simi, E., Moyer, K., Aufox, S., Ouyang, D.W., 2014. Clinical implementation of noninvasive prenatal testing among maternal fetal medicine specialists. Prenat. Diagn. 34 (5), 416–423.

Juneau, K., Bogard, P.E., Huang, S., Mohseni, M., Wang, E.T., Ryvkin, P., Kingsley, C., Struble, C.A., Oliphant, A., Zahn, J.M., 2014. Microarray-based cell-free DNA analysis improves noninvasive prenatal testing. Fetal Diagn. Ther. 36 (4), 282–286.

Kalousek, D.K., Barrett, I.J., McGillivray, B.C., 1989. Placental mosaicism and interuterine survival of trisomies 13 and 18. Am. J. Hum. Genet. 44, 338–343.

Konialis, C., Pangalos, C., 2015. Dilemmas in prenatal chromosomal diagnosis revealed through a single center's 30 years' experience and 90,000 cases. Fetal Diagn. Ther. 38, 218–232.

Koumbaris, G., Kypri, E., Tsangaras, K., Achilleos, A., Mina, P., Neofytou, M., Velissariou, V., Christopoulou, G., Kallikas, I., González-Liñán, A., Benusiene, E., Latos-Bielenska, A., Marek, P.,

Santana, A., Nagy, N., Széll, M., Laudanski, P., Papageorgiou, E.A., Ioannides, M., Patsalis, P.C., 2016. Cell-free DNA analysis of targeted genomic regions in maternal plasma for non-invasive prenatal testing of trisomy 21, trisomy 18, trisomy 13 and fetal sex. Clin. Chem. 62, 848–855.

Kwok, P.Y., Gu, Z., 1999. Single nucleotide polymorphism libraries: why and how are we building them? Mol. Med. Today 5 (12), 538–543. Review.

Lau, T.K., Cheung, S.W., Lo, P.S.S., Pursley, A.N., Chan, M.K., Jiang, F., Zhang, H., Wang, W., Jong, L.F.J., Yuen, O.K.C., Chan, H.Y.C., Chan, W.S.K., Choy, K.W., 2014. Non-invasive prenatal testing for fetal chromosomal abnormalities by low-coverage whole-genome sequencing of maternal plasma DNA: review of 1982 consecutive cases in a single center. Ultrasound Obstet. Gynecol. 43, 254–264.

Liang, D., Lv, W., Wang, H., Xu, L., Liu, J., Li, H., Hu, L., Peng, Y., Wu, L., 2013. Non-invasive prenatal testing of fetal whole chromosome aneuploidy by massively parallel sequencing. Prenat. Diagn. 33, 409–415.

Liao, G.J., Chan, K.C., Jiang, P., Sun, H., Leung, T.Y., Chiu, R.W., Lo, Y.M., 2012. Noninvasive prenatal diagnosis of fetal trisomy 21 by allelic ratio analysis using targeted massively parallel sequencing of maternal plasma DNA. PLoS One 7, e38154.

Litton, C., Stone, J., Eddleman, K., Lee, M.J., 2009. Noninvasive prenatal diagnosis: past, present, and future. Mt. Sinai J. Med. 76, 521–528.

Lo, Y.M.D., Corbetta, N., Chamberlain, P.F., Rai, V., Sargent, I.L., Redman, C.W., et al., 1997. Presence of fetal DNA in maternal plasma and serum. Lancet 350485–350487.

Lo, Y.M.D., Tein, M.S., Lau, T.K., Haine, C.J., Leung, T.N., Poon, P.M., Wainscoat, J., Johnson, P.J., Chang, A.M., Hjelm, N.M., 1998a. Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. Am. J. Hum. Genet. 62, 768–775.

Lo, Y.M.D., Hjelm, N.M., Fidler, C., Sargent, I.L., Murphy, M.F., Chamberlain, P.F., Poon, P.M., Redman, C.W., Wainscoat, J.S., 1998b. Prenatal diagnosis of fetal RhD status by molecular analysis of maternal plasma. N. Engl. J. Med. 339, 1734–1738.

Lo, Y.M.D., Lau, T.K., Zhang, J., Leung, T.N., Chang, A.M., Hjelm, N.M., Elmes, R.S., Bianchi, D.W., 1999a. Increased fetal DNA concentrations in the plasma of pregnant women carrying fetuses with trisomy 21. Clin. Chem. 45, 1747–1751.

Lo, Y.M.D., Zhang, J., Leung, T.N., Lau, T.K., Chang, A.M., Hjelm, N.M., 1999b. Rapid clearance of fetal DNA from maternal plasma. Am. J. Hum. Genet. 64, 218–224.

Lo, Y.M.D., Lun, F.M., Chan, K.C., Tsui, N.B., Chong, K.C., Lau, T.K., Leung, T.Y., Zee, B.C., Cantor, C.R., Chiu, R.W., 2007a. Digital PCR for the molecular detection of fetal chromosomal aneuploidy. Proc. Natl. Acad. Sci. U.S.A. 104, 13116–13121.

Lo, Y.M.D., Tsui, N.B., Chiu, R.W., Lau, T.K., Leung, T.N., Heung, M.M., Gerovassili, A., Jin, Y., Nicolaides, K.H., Cantor, C.R., Ding, C., 2007b. Plasma placental RNA allelic ratio permits noninvasive prenatal chromosomal aneuploidy detection. Nat. Med. 13, 218–223.

Lo, Y.M.D., Chan, K.C., Sun, H., Chen, E.Z., Jiang, P., Lun, F.M., Zheng, Y.W., Leung, T.Y., Lau, T.K., Cantor, C.R., Chiu, R.W., 2010. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. Sci. Transl. Med. 2, 61 ra91.

Lo, Y.M.D., 2013. Non-invasive prenatal testing using massively parallel sequencing of maternal plasma DNA: from molecular karyotyping to fetal whole-genome sequencing. Rev. Reprod. Biomed. Online 27 (6), 593–598.

Lui, Y.Y., Chik, K.W., Chiu, R.W., Ho, C.Y., La, C.W., Lo, Y.M.D., 2002. Predominant hematopoietic origin of cell-free DNA in plasma and serum after sex-mismatched bone marrow transplantation. Clin. Chem. 48, 421–427.

Mann, K., Hills, A., Donaghue, C., Thomas, H., Ogilvie, C.M., 2012. Quantitative fluorescence PCR analysis of >40,000 prenatal samples for the rapid diagnosis of trisomies 13, 18 and 21 and monosomy X. Prenat. Diagn. 32, 1197–1204.

Masuzaki, H.K., Yoshiura, K.I., Yoshimur, S., Niikawa, N., Ishimaru, T., 2004. Detection of cell free placental DNA in maternal plasma: direct evidence from three cases of confined placental mosaicism. J. Med. Genet. 41, 289–292.

Mennuti, M.T., Cherry, A.M., Morrissette, J.J., Dugoff, L., 2013. Is it time to sound an alarm about false-positive cell-free DNA testing for fetal aneuploidy? Am. J. Obstet. Gynecol. 209 (5), 415–419.

Monk, M., 1990. Changes in DNA methylation during mouse embryonic development in relation to X-chromosome activity and imprinting. Philos. Trans. R. Soc. Lond. 326, 299–312.

Morris, A., Boyd, E., Dhanjal, S., Lowthe, G.W., Aitken, D.A., Youn, J., Menzies, A.L., Imrie, S.J., Connor, J.M., 1999. Two years' prospective experience using fluorescence in situ hybridization on uncultured amniotic fluid cells for rapid prenatal diagnosis of common chromosomal aneuploidies. Prenat. Diagn. 19, 546–551.

Nicolaides, K.H., Syngelaki, A., Ashoor, G., Birdir, C., Touzet, G., 2012. Noninvasive prenatal testing for fetal trisomies in a routinely screened first-trimester population. Am. J. Obstet. Gynecol. 207 (5), 374.

Nicolaides, K.H., Syngelaki, A., Gil, M., Atanasova, V., Markova, D., 2013. Validation of targeted sequencing of single-nucleotide polymorphisms for non-invasive prenatal detection of aneuploidy of chromosomes 13, 18, 21, X and Y. Prenat. Diagn. 33, 575–579.

Nicolaides, K.H., 2011. Turning the pyramid of prenatal care. Fetal Diagn. Ther. 29, 183–196.

Norton, M.E., Brar, H., Weiss, J., Karim, I.A., Laurent, L.C., Caughey, A.B., Rodriguez, M.H., Williams 3rd, J., Mitchell, M.E., Adair, C.D., Lee, H., Jacobsson, B., Tomlinson, M.W., Oepkes, D., Hollemon, D., Sparks, A.B., Oliphant, A., Song, K., 2012. Non-invasive chromosomal evaluation (NICE) study: results of a multicenter prospective cohort study for detection of fetal trisomy 21 and trisomy 18. Am. J. Obstet. Gynecol. 207, 137.e1-8.

Nygren, A.O., Dean, J., Jensen, T.J., Kruse, S., Kwong, W., van den Boom, D., Ehrich, M., 2010. Quantification of fetal DNA by use of methylation-based DNA discrimination. Clin. Chem. 56, 1627–1635.

Osborne, C.M., Hardisty, E., Devers, P., Kaiser-Rogers, K., Hayden, M.A., Goodnight, W., Vora, N.L., 2013. Discordant noninvasive prenatal testing results in a patient subsequently diagnosed with metastatic disease. Prenat. Diagn. 33, 609–611.

Palomaki, G.E., Kloza, E.M., Lambert-Messerlian, G.M., Haddow, J.E., Neveux, L.M., Ehrich, M., van den Boom, D., Bombard, A.T., Deciu, C., Grody, W.W., Nelson, S.F., Canick, J.A., 2011. DNA sequencing of maternal plasma to detect Down syndrome: an international clinical validation study. Genet. Med. 13, 913–920.

Palomaki, G.E., Deciu, C., Kloza, E.M., Lambert-Messerlian, G.M., Haddow, J.E., Neveux, L.M., Ehrich, M., van den Boom, D., Bombard, A.T., Grody, W.W., Nelson, S.F., Canick, J.A., 2012. DNA sequencing of maternal plasma reliably identifies trisomy 18 and

trisomy 13 as well as Down syndrome: an international collaborative study. Genet. Med. 14 (3), 296−305.

Papageorgiou, E.A., Fiegler, H., Rakyan, V., Beck, S., Hulten, M., Lamnissou, K., Carter, N.P., Patsalis, P.C., 2009. Sites of differential DNA methylation between placenta and peripheral blood: molecular markers for noninvasive prenatal diagnosis of aneuploidies. Am. J. Pathol. 174, 1609−1618.

Papageorgiou, E.A., Karagrigoriou, A., Tsaliki, E., Velissariou, V., Carter, N.P., Patsalis, P.C., 2011. Fetal specific DNA methylation ratio permits non-invasive prenatal diagnosis of trisomy 21. Nat. Med. 17, 510−513.

Papageorgiou, E.A., Koumbaris, G., Kypri, E., Hadjidaniel, M., Patsalis, P.C., 2014. The epigenome view: an effort towards non-invasive prenatal diagnosis. Rev. Genes 5, 310−329.

Patsalis, P.C., Tsaliki, E., Koumbaris, G., Karagrigoriou, A., Velissariou, V., Papageorgiou, E., 2012. A new non-invasive prenatal diagnosis of Down syndrome through epigenetic markers and real-time qPCR. Expert Opin. Biol. Ther. (Suppl. 1), S155−S161.

Petersen, O.B., Vogel, I., Ekelund, C., Hyett, J., Tabor, A., 2014. Danish Fetal Medicine Study Group; Danish Clinical Genetics Study Group. Potential diagnostic consequences of applying non-invasive prenatal testing: population-based study from a country with existing first-trimester screening. Ultrasound Obstet. Gynecol. 43 (3), 265−271.

Poon, L.L.M., Leung, T.N., Lau, T.K., Chow, K.C.K., Lo, Y.M.D., 2002. Differential DNA methylation between fetus and mother as a strategy for detecting fetal DNA in maternal plasma. Clin. Chem. 48, 35−41.

Quezada, M.S., Gil, M.M., Francisco, C., Orosz, G., Nicolaides, K.H., 2015. Screening for trisomies 21, 18 and 13 by cell-free DNA analysis of maternal blood at 10−11 weeks' gestation and the combined test at 11−13 weeks. Ultrasound Obstet. Gynecol. 45, 36−41.

Rava, R.P., Srinivasan, A., Sehnert, A.J., Bianchi, D.W., 2013. Circulating fetal cell-free DNA fractions differ in autosomal aneuploidies and monosomy X. Clin. Chem. 60, 243−250.

Reik, W., Dean, W., Walter, J., 2001. Epigenetic reprogramming in mammalian development. Science 293, 1089−1093.

Sagoo, G.S., Butterworth, A.S., Sanderson, S., Shaw-Smith, C., Higgins, J.P., Burton, H., 2009. Array CGH in patients with learning disability (mental retardation) and congenital anomalies: updated systematic review and meta-analysis of 19 studies and 13,926 subjects. Genet. Med. 11, 139−146.

Schübeler, D., 2015. ESCI award lecture: regulation, function and biomarker potential of DNA methylation. Eur. J. Clin. Invest. 45, 288−293.

Schuring-Blom, G.H., Boer, K., Leschot, N.J., 2001. A placental diploid cell line is not essential for ongoing trisomy 13 or 18 pregnancies. Eur. J. Hum. Genet. 9, 286−290.

Sehnert, A.J., Rhees, B., Comstock, D., Comstock, D., de Feo, E., Heilek, G., Burke, J., Rava, R.P., 2011. Optimal detection of fetal chromosomal abnormalities by massively parallel DNA sequencing of cell-free fetal DNA from maternal blood. Clin. Chem. 57, 1042−1049.

Sifakis, S., Koukou, Z., Spandidos, D.A., 2015. Cell-free fetal DNA and pregnancy-related complications (review). Mol. Med. Rep. 11, 2367−2372.

Smith, M., Lewis, K.M., Holmes, A., Visootsak, J., 2014. A case of false negative NIPT for Down syndrome − lessons learned. Case Rep. Genet. 2014, 823504.

Sparks, A.B., Wang, E.T., Struble, C.A., Barrett, W., Stokowski, R., McBride, C., Zahn, J., Lee, K., Shen, N., Doshi, J., Sun, M., Garrison, J., Sandler, J., Hollemon, D., Pattee, P., Tomita-Mitchell, A., Mitchell, M., Stuelpnagel, J., Song, K., Oliphant, A., 2012. Selective analysis of cell-free DNA in maternal blood for evaluation of fetal trisomy. Prenat. Diagn. 32, 3−9.

Srinivasan, A., Bianchi, D.W., Liao, W., Sehnert, A.J., Rava, R.P., 2013. Maternal plasma DNA sequencing: effects of multiple gestation on aneuploidy detection and the relative cell-free fetal DNA (cffDNA) per fetus. Am. J. Obstet. Gynecol. 208, S31.

Tabor, A., Alfirevic, Z., 2010. Update on procedure-related risks for prenatal diagnosis techniques. Fetal Diagn. Ther. 27, 1−7.

Taylor-Phillips, S., Freeman, K., Geppert, J., Agbebiyi, A., Uthman, O.A., Madan, J., Clarke, A., Quenby, S., Clarke, A., January 18, 2016. Accuracy of non-invasive prenatal testing using cell-free DNA for detection of Down, Edwards and Patau syndromes: a systematic review and meta-analysis. BMJ Open 6 (1).

Tounta, G., Vrettou, C., Kolialex, A., Papantoniou, N., Destouni, A., Tsangaris, G.T., Antsaklis, A., Kanavakis, E., Mavrou, A., 2011. A multiplex PCR for non-invasive fetal RHD genotyping using cell-free fetal DNA. Vivo 25, 411−417.

Tsaliki, E., Papageorgiou, E.A., Spyrou, C., Koumbaris, G., Kypri, E., Kyriakou, S., Sotiriou, C., Touvana, E., Keravnou, A., Karagrigoriou, A., Lamnissou, K., Velissariou, V., Patsalis, P.C., 2012. MeDIP real-time qPCR of maternal peripheral blood reliably identifies trisomy 21. Prenat. Diagn. 32, 996−1001.

Tsui, N.B., Jiang, P., Wong, Y.F., Leung, T.Y., Chan, K.C., Chiu, R.W., Sun, H., Lo, Y.M., 2014. Maternal plasma RNA sequencing for genome-wide transcriptomic profiling and identification of pregnancy-associated transcripts. Clin. Chem. 60, 954−962.

Verweij, E.J., de Boer, M.A., Oepkes, D., 2014. Non-invasive prenatal testing for trisomy 13: more harm than good? Ultrasound Obstet. Gynecol. 44, 112−114.

Vora, N.L., Johnson, K.L., Basu, S., Catalano, P.M., Hauguel-de Mouzon, S., Bianchi, D.W., 2012. A multifactorial relationship exists between total circulating cell-free DNA levels and maternal BMI. Prenat. Diagn. 32, 912−914.

Wallerstein, R., Yu, M., Neu, R.L., Benn, P., Lee Bowen, C., Crandall, B., Disteche, C., Donahue, R., Harrison, B., Hershey, D., Higgins, R.R., Jenkins, L.S., Jackson-Cook, C., Keitges, E., Khodr, G., Lin, C.C., Luthardt, F.W., Meisner, L., Mengden, G., Patil, S.R., Rodriguez, M., Sciorra, L.J., Shaffer, L.G., Stetten, G., Van Dyke, D.L., Wang, H., 2000. Common trisomy mosaicism diagnosed in amniocytes involving chromosomes 13, 18, 20, and 21: karyotype-phenotype correlations. Prenat. Diagn. 20, 103−122.

Wang, E., Batey, A., Struble, C., Musci, T., Song, K., Oliphant, A., 2013. Gestational age and maternal weight effects on fetal cell-free DNA in maternal plasma. Prenat. Diagn. 33, 662−666.

Wapner, R.J., Babiarz, J.E., Levy, B., Stosic, M., Zimmermann, B., Sigurjonsson, S., Wayham, N., Ryan, A., Banjevic, M., Jing Hu, L., Hall, M.P., Demko, Z., Siddiqui, A., Rabinowitz, M., Gross, S.J., Hill, M., Benn, P., 2015. Expanding the scope of noninvasive prenatal testing: detection of fetal microdeletion syndromes. Am. J. Obstet. Gynecol. 212 (3), 332.e1-9.

Weiss, K.M., 1998. In search of human variation. Genome Res. 8, 691−697. Review.

Wright, D., Syngelaki, A., Bradbury, I., Akolekar, R., Nicolaides, K.H., 2014. First-trimester screening for trisomies 21, 18 and 13 by ultrasound and biochemical testing. Fetal Diagn. Ther. 35, 118–126.

Wright, D., Wright, A., Nicolaides, K.H., 2015. A unified approach to risk assessment for fetal aneuploidies. Ultrasound Obstet. Gynecol. 45, 48–54.

Yao, H., Zhang, L., Zhang, H., Jiang, F., Hu, H., Chen, F., Jiang, H., Mu, F., Zhao, L., Liang, Z., Wang, W., 2012. Noninvasive prenatal genetic testing for fetal aneuploidy detects maternal trisomy X. Prenat. Diagn. 32, 1114–1116.

Zhang, H., Gao, Y., Jiang, F., et al., 2015. Noninvasive prenatal testing for trisomy 21, 18 and 13 – clinical experience from 146,958 pregnancies. Ultrasound Obstet. Gynecol. 45 (5), 530–538.

Zimmermann, B., Hill, M., Gemel, G., Demko, Z., Banjevi, M., Bane, J., Ryan, A., Sigurjonsson, S., Chopra, N., Dodd, M., Levy, B., Rabinowitz, M., 2012. Noninvasive prenatal aneuploidy testing of chromosomes 13, 18, 21, X, and Y, using targeted sequencing of polymorphic loci. Prenat. Diagn. 32, 1233–1241.

Chapter 25

Genetic Testing and Psychology

S.R. Brand and A.F. Patenaude
Dana-Farber Cancer Institute, Boston, MA, United States

25.1 INTRODUCTION

The Human Genome Project (HGP), completed in 2003, provided the first comprehensive sequence of a human genome. This huge international cooperative effort produced a tool of enormous importance to basic scientists, clinical researchers, clinicians, and, ultimately, patients. A great deal has been learned about basic biology from the HGP, including the fact that humans have only about 20,000 protein-coding genes, many fewer than originally estimated. However, the most interesting challenges lie ahead (Feero et al., 2008). With the human genome sequence readily available to all, the challenge is to "capitalize on the immense potential of the HGP to improve human health and well-being" (Collins et al., 2003).

Genetic advances are continuing to change the practice of medicine in significant ways. Next-generation sequencing technologies, which include whole-genome and exome sequencing, are becoming more available in hospitals and laboratories throughout the world. The costs associated with genetic testing and even genome sequencing have dropped dramatically (Phimister et al., 2012). Genomic medicine is increasingly being integrated into primary care, increasing its accessibility dramatically (David et al., 2015).

In 2003, the United States National Human Genome Research Institute put forth a blueprint for future genomic research. The second of three major challenges, "Genomics to Health," involved the translation of "genome-based knowledge into health benefits." The goals were to:

- "Identify genes and pathways with a role in health and disease, and determine how they interact with environmental factors,
- Develop, evaluate, and apply genome-based diagnostic methods for the prediction of susceptibility to disease, the prediction of drug response, the early detection of illness, and the accurate molecular classification of disease, and
- Develop and deploy methods that catalyze the translation of genomic information into therapeutic advances" (Collins et al., 2003).

It is argued that the achievement of these lofty aims is dependent on psychological factors in a variety of ways. Whether patients are willing to participate in genetic research, whether they opt to undergo genetic counseling or testing to determine if they have an increased disease risk, whether they are willing to adopt screening or surveillance recommendations targeted to those known or suspected of being at high risk, and whether they opt for prophylactic surgery depend in large measure on emotional, attitudinal, and behavioral factors. The recruitment of appropriate, diverse patients for genetic research studies will necessitate an improved understanding of what patients bring to genetic counseling or testing from their own family experience with illness, their perception of disease risk, and their attitudes toward genetics research and the medical community. Individuals from different ethnic and cultural communities vary widely in their attitudes toward genetics, genetic testing, and genetic medicine, and more information is needed to optimally counsel an increasingly diverse patient population in cancer genetics.

Subjective feelings play major roles in determining how people respond to situations they consider threatening. Research suggests that individuals often act on the basis of subjective feelings about their disease risk, even after receiving counseling that informs them that their risk perceptions are overestimates of actual risk (Iglehart et al., 1998). How can such responses be modified to be in tune with actual risk? What kinds of counseling techniques need to be developed to guide patients toward selecting one of the appropriate choices for mutation carriers? How can people who are found to be at an increased hereditary risk

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00025-0
Copyright © 2017 Elsevier Ltd. All rights reserved.

but who do not heed surveillance or screening recommendations be educated? How can people be helped to make optimal personal decisions about prophylactic surgery, often the risk-reduction option that conveys the greatest protection against developing disease but at a significant emotional cost? What is known about how individuals in high-risk families communicate about hereditary disease predisposition with their relatives?

Much of the ultimate success of the HGP in improving human health rests on whether individuals, suspected or known to be at risk for hereditary illness, feel sufficiently comfortable with the use of genetic technology and with the emotional impact of genetic information to make use of the revolutionary advances that the HGP has made possible. This chapter attempts to summarize what is currently known about psychological factors as they affect individuals and families with concerns about hereditary predispositions to illness.

With the advent of genetic testing, first for Huntington's disease (HD) and over the past several decades for several types of cancer, we have been able to observe reactions to the increased availability of genetic information in individuals from hereditary disease families. Going forward, there will increasingly be studies of the psychosocial impact of genetic counseling and testing for a broadening range of disease genes. In the United States and in several other countries, studies of the ethical, legal, social, and psychological impacts of genetic testing have been undertaken simultaneously with molecular studies of the human genome. It is critical to learn under what circumstances genetic advances improve people's lives and health as well as understanding under what circumstances adverse outcomes result from participation in genetic research, genetic counseling, and/or genetic testing.

25.2 GETTING TO THE TEST: AWARENESS, ACCESS, AND ADVERTISING

25.2.1 What Tests Are There?

One major change involves the clinical uptake of panel testing (Slavin et al., 2015). An increasing number of mutations have been identified which predispose to common cancers, making panel testing possible and, in many cases, efficient and cost-effective. For example, with panel testing, a woman concerned about her hereditary risk for breast cancer might be tested not only for the breast/ovarian cancer gene mutations that were first identified but also for several dozen or even more genes with lower penetrance (Easton et al., 2015) (e.g., *CDH1*, which is associated with breast, gastric, and colorectal cancer; *PTEN*, which causes breast, thyroid, uterine, colorectal, and kidney cancer; and *STK11*, which is associated with colorectal, breast, small bowel, pancreatic, and ovarian cancers). Panel tests replace the only previous option (sequential genetic testing) to find the underlying cause of a hereditary predisposition disorder. Panels may be specific for genes that are associated with breast cancer or Lynch syndrome or for a range of pediatric cancer genes. Panel testing may be helpful in identifying genetic mutations that were not clearly indicated by family history. They can, however, greatly increase the element of surprise in the receipt of test findings in comparison with single gene testing. The psychological impact of informing individuals about carrying hereditary predisposing mutations that were not among those previously considered or considered likely is still being evaluated. Because there are so many genes on a single panel, it is difficult in pretest counseling to seriously consider and discuss the whole range of results that may be found and their medical and psychological implications. As panel testing is increasingly in use, more work is needed to identify the psychological risks of such testing and to train genetic counselors and others to help patients consider the impact of unanticipated findings that may result from panel testing.

Similarly, genomic sequencing is increasingly being incorporated into clinical medicine and carries with it similar risks of unanticipated and potentially unwanted findings. Various professional bodies have identified genetic mutations that are clearly deleterious and actionable, and advice has been offered about which findings need to be shared with tested individuals, if they are willing to receive this information (Robson et al., 2015). With genomic sequencing, there is also a risk that mutations may be found that have major unexpected or unanticipated implications for relatives, sometimes more so than they do for the tested individual, so that sharing of the information with family members is strongly advised. Genomic sequencing also makes known genetic mutations that predispose to adult-onset cancers, for which children are not traditionally tested, often with complicated consequences. For example, it may be that sequencing of the tumor of a 6-year-old pediatric cancer patient reveals a mutation in a *BRCA1* mutation, which is subsequently also seen on the 6-year-old patient's germline sequencing. In this case, it would be critical that the parents be informed and the mother tested to see if she is also a carrier of a *BRCA1/2* mutation to learn if she is at a significantly higher than average risk of both breast and ovarian cancer. Hence while the power of genetic and genomic tests has increased remarkably, so, too, have the ethical and psychological issues that should be considered.

25.2.2 Genetic Testing

Genetic testing was first performed in research settings and offered to members of families known to have had unusual

numbers of affected individuals (i.e., individuals diagnosed with a disease or diseases of hereditary etiology). Members of these families tended to have been aware for a long time, sometimes several generations, about the unusual concentration of disease in their families. Many were from families with extremely high concentrations of hereditary cancer, since it was the most affected families who first came to the notice of clinicians and disease registries. Family members had often observed multiple early deaths from the disease in their family and had grown up fearing that this would be their fate as well.

Cancer genetic testing has moved beyond the research setting into clinical testing, with some advising population screening for *BRCA1/2* (Manchanda et al., 2015; Lu, 2015). Clinical testing makes it likely that individuals with less extensive family histories of disease will be tested and also more likely that genetic testing occurs in closer proximity to the cancer diagnosis. The impact of this change has not yet been fully evaluated, but research to date suggests that patients view such testing favorably (Wevers et al., 2012; Zilliacus et al., 2012), and major, long-term psychological problems have not been reported to date (Manchanda et al., 2015). Genetic testing at diagnosis increases the uptake of prophylactic mastectomy among mutation carriers (Weitzel et al., 2003; Schwartz et al., 2004). Patients who elect contralateral prophylactic mastectomy (CPM) after "upfront genetic testing" do not experience more distress than patients who do not undergo CPM in the year following a breast cancer diagnosis (Tercyak et al., 2007).

Some companies offer patients direct access to test kits via the Internet (Williams-Jones, 2003), which raises significant issues among professional groups and ethicists about assuring that individuals undergoing testing are sufficiently educated about the implications of learning their test results and that the testing offers sufficient privacy protection (Edwards and Huang, 2014; Geransar and Einsiedel, 2008). Changes in physician behavior may result from increased patient requests as a result of the advertising of these "direct to consumer" tests (Myers et al., 2006).

Genetic testing does not supply all the answers patients wish to have. There are very real limitations to what genetic testing can tell patients about the likelihood they will develop the disease in question. A positive test result for a gene that is less than 100% penetrant does not indicate for certain that the mutation carrier will ever develop the disease or when it will occur. In some cases, a single genetic mutation predisposes the individual to a number of related diseases; testing usually will not indicate which disease the person will develop or whether they will develop more than one disease within the syndrome of diseases to which they are predisposed. Cancer genetic testing is, after all, a trade of the uncertainty about whether one is a mutation carrier for the uncertainty, in the event that the test is positive, about which cancers might appear and when they will emerge.

25.2.3 Access to Testing

Access to genetic testing is not equally available to all. Full-sequence testing for cancer genes costs several thousand dollars, and patients may opt not to bill the costs to their health insurance, either because it is not a covered service in all plans or because patients are reluctant to share test results with their insurer for fear of later discrimination. In the United States, passage of the Genetic Information Nondiscrimination Act prevents health insurers from rejecting healthy individuals or raising health insurance premiums on the basis of personal or familial genetic predispositions. It does not, however, cover life, disability, or long-term care insurance (Hudson et al., 2008). In countries with a national health insurance system, discrimination is not an issue for health insurance. However, access to genetic services may be limited by rationing of the number of genetic tests offered in a geographic region during a particular time period and/or very strict criteria for test eligibility. These practices may result in long waits for testing or may exclude patients with less extensive family histories.

Minority populations are likely to be particularly hindered in their access to genetic testing for a variety of reasons, including cost, a lack of minority genetics professionals, and a lack of non-English language educational materials. A 2011 review (Levy et al., 2011) of a US national sample of 14.4 million commercially insured patients showed marked discrepancies in the genetic testing experience of 1474 young women (ages 20–40) diagnosed with early-onset breast cancer. Understandably, given the well-known founder mutations for *BRCA1/2*, Jewish women of Ashkenazi descent were nearly three times more likely to receive genetic testing than non-Jewish women in this cohort, and women of color were much less likely to be tested. The hazard ratios were 0.34 for black women and 0.52 for Hispanic women. The type of insurance also affected testing rates; women in health maintenance organizations were significantly less likely to get testing than women in point of service plans.

25.3 INDIVIDUAL FACTORS INFLUENCING THE UTILIZATION OF GENETIC TESTING

Complex psychological factors govern an individual's uptake of genetic testing and the use he or she makes of the information received. Personal experience with a family member who has had the disease in question and possibly died from it is likely to impact one's perception of personal risk, disease-related anxiety, and the desire to avoid a similar outcome. Other factors, including general personality characteristics, can affect how actively a person seeks to confront knowledge of personal, hereditary disease risk

and to undertake actions designed to prevent or detect illness. Increasing knowledge of these factors enables researchers and counselors to develop more effective genetic counseling and follow-up techniques and to better advise both patients and medical providers.

25.3.1 Risk Perception

Members of high-risk cancer families typically greatly overestimate their risk for developing cancer (van Oostrom et al., 2006; Kash et al., 2000). Those who have lost a parent to the disease are especially likely to overestimate the risk that they, themselves, will develop the cancer their parent died from. In an English study of women referred for genetic counseling to a Cancer Family History Clinic, 70% could not cite the population risk for breast cancer, but 76% perceived themselves to be at higher than population levels of risk for breast cancer (Hallowell et al., 1998). A study by Zakowski et al. (1997) found that women undergoing mammography screening (but with no abnormal findings) whose mother had died of breast cancer estimated their lifetime risk of developing breast cancer as 70% versus the 53% risk estimate of women undergoing mammography who did not have a mother die of breast cancer and the 32% risk estimate of similarly aged women in the general population not undergoing mammography studies. All of these figures significantly overestimate breast cancer risk; the lifetime risk for breast cancer for a woman in the United States general population is about 10%. Having a first-degree relative with breast cancer raises that risk from 1.7 to 5 times (Offit and Geffen, 1998) but not to 70%.

Similarly, the experience of deaths of close relatives due to an hereditary nonpolyposis colorectal cancer (HNPCC)-related cancer increased risk perception among HNPCC family members (Domanska et al., 2007). In families with an identified mutation in *BRCA1/2* or an HNPCC-related gene, the age at which a parent was diagnosed with cancer had a significant effect on offspring risk perception in both groups. Individuals who, in their elementary school years or earlier experienced having a parent with cancer, reported higher levels of perceived risk compared to those whose parents had cancer during their adolescent or adult years (van Oostrom et al., 2006).

A high perceived risk of cancer is a source of motivation for seeking genetic counseling or testing, although high-risk perception may also be accompanied by anxiety, which deters some individuals from accepting counseling or testing (McInerney-Leo et al., 2006). Genetic testing provides objective information about a cancer risk and is, in most studies, an important factor affecting risk perception. Given that many women come to *BRCA1/2* genetic testing with high perceptions of their own breast cancer risk, the receipt of a positive test result indicating that one is a carrier may not significantly raise risk perception (McInerney-Leo et al., 2006). Young women who are carriers are particularly likely to have their risk perception increase following genetic testing that shows them to be carriers, but the level of risk they perceive may decrease with time (Watson et al., 2005). For *BRCA1/2* mutation carriers, perception of their risk for ovarian cancer is likely to increase following the receipt of a positive test result (McInerney-Leo et al., 2006; Cherry et al., 2013). Those found to be negative for a known familial mutation are likely to have decreased risk perception (Watson et al., 2005; Schwartz et al., 2002; Claes et al., 2005). Interestingly, for *BRCA1/2* mutation carriers, lower perceived risk prior to testing predicted lower anxiety scores after receiving positive results (Bredart et al., 2013).

Similarly, perceptions of risk among carriers of genes conveying an increased predisposition to colon cancer increase following the disclosure of positive test results, but there are even larger, significant increases among carriers following result disclosure in perception of their risk for extracolonic cancers (Hadley et al., 2008). One review found that while multiple studies have demonstrated an improvement in the accuracy of risk perception after genetic testing, such improvement is not universal (Nelson et al., 2014). Genetic counseling and testing are particularly important, then, in notifying members of high-risk cancer families about their risk for cancers other than those that have primarily affected family members previously. It is important to note that the relationship between perceived risk and risk management decisions is not always direct (Hoskins et al., 2012). Multiple studies have suggested that it is the counselee's recollection and interpretation of the information presented by genetic counselors, as opposed to the actual genetic information provided, that best predicts behavioral outcomes of genetic testing (Vos et al., 2011, 2012a,b).

25.3.2 Distress

Genetic counseling has not been shown to increase distress in high-risk individuals (Nelson et al., 2014; Bleiker et al., 2013). Multiple studies have described the impact of a positive *BRCA1/2* testing result in the year following the testing, with most showing that *BRCA1/2* mutation carriers experience higher levels of distress compared to the noncarriers in the first months following testing but that differences attenuate over the first year (Hamilton et al., 2009). Similar patterns have been found in individuals undergoing genetic testing for Lynch syndrome (Bleiker et al., 2013).

However, only a few studies have examined the trajectory of distress for longer than the first year, and more longitudinal studies are needed to fully understand the impact of testing on distress over time. A longitudinal study prospectively examining psychosocial outcomes following *BRCA1/2* testing found that 5 years after testing, women

who were mutation carriers continued to report higher levels of genetic testing distress than women who were not carriers (Graves et al., 2012). These findings may reflect the family disruptions that can occur in relation to hereditary disease that may continue to cause distress even after the initial shock of diagnosis has dissipated and suggest a need for ongoing psychosocial support in these families of note, the use of specific measures of genetic testing distress may be more sensitive to salient aspects of distress for a particular population, compared to more global measures of distress used in other studies (Cella et al., 2002).

In many studies, there is a subgroup of tested individuals who do experience significant disease or testing-related distress. The experience of cancer in a close relative has been linked to higher levels of distress, with some participants indicating distress specifically linked to painful memories of family members who had the familial disease (Turner-Cobb et al., 2006; Butow et al., 2005). Having a parent with cancer during one's childhood is associated with an increased risk for psychological distress when undergoing cancer genetic testing as an adult (van Oostrom et al., 2007). Other risk factors for increased psychological distress among individuals undergoing predictive genetic counseling have been identified and include being female (Vernon et al., 1997), having high levels of distress prior to counseling (Murakami et al., 2004), a lack of social support (Lammens et al., 2010), and having a previous cancer diagnosis (Douma et al., 2010).

Early work has shown that cognitive emotional factors can also be related to distress around hereditary disease risk. Miller et al. (1995) distinguished two styles of coping with threatening information. *Monitors* are those who tend to focus on the threat, to actively seek information related to the threat, and to have difficulty perceiving that all has been done that could be done to cope with the threat. *Blunters*, in contrast, are individuals who naturally try to avoid or shield themselves from threatening information. It has been shown that anxiety during the wait for genetic testing results is higher among women who are high monitors (Tercyak et al., 2002), suggesting that these women may need extra support while awaiting their results. Similarly, an Israeli study of individuals undergoing colon cancer genetic testing found that monitors were more distressed than nonmonitors, especially when they received indeterminate or positive results (Shiloh et al., 2008). In addition, one review found that monitors tended to be more demanding of their healthcare providers in terms of the desire for information and emotional support (Roussi and Miller, 2014). Cognitive style also appears to influence participation in screening; the targeting of screening and surveillance recommendations according to the patient's cognitive style may be important to maximize participation (Miller et al., 1995; Roussi et al., 2011).

25.3.3 Health Beliefs

Misconceptions can affect one's ability to realistically evaluate the increase in risk imposed by being a mutation carrier and can also affect one's thinking about the value of screening, surveillance, or prophylactic surgery options. There are many misconceptions about inherited cancer risk, even among those at increased risk. For example, daughters of living *BRCA1/2*-positive mothers have been shown to have suboptimal genetic knowledge with significant misconceptions, including when it is possible to undergo genetic testing, which cancers are associated with *BRCA1/2* mutations, and whether breast cancer in *BRCA1/2* mutation carriers occurs at earlier ages than breast cancers in the general population (Patenaude et al., 2013a).

It has been shown that when high-profile celebrities disclose health information, there is a significant impact on the public's awareness and understanding of that disease or condition. Angelina Jolie's announcement in the *New York Times* in 2013 that she had undergone a prophylactic double mastectomy after testing positive for *BRCA1* (Jolie, 2013) led to a dramatic increase in online searches about breast cancer and genetic testing (Noar et al., 2015), including a 795-fold increase in views of the preventative mastectomy fact sheet on the National Cancer Institute's webpage (Juthe et al., 2015). At least one study demonstrated a sustained effect of her public disclosure on referrals for genetic counseling and testing. Evans and colleagues examined referral data to breast cancer family history clinics and regional genetics centers in the United Kingdom and found an almost 2.5-fold increase in referrals in the first 2 months following Jolie's announcement. In addition, requests for *BRCA1/2* testing almost doubled during this period, with no apparent increase in the number of inappropriate referrals (Evans et al., 2014).

Different ethnic or cultural groups may have different beliefs and hold different values related to medicine generally and to genetics more specifically. African Americans, Hispanics, and Asians often are found to have lower perceived cancer risks than Whites (Davis et al., 2013; Orom et al., 2010) and are less willing to undergo molecular testing for personalized cancer therapy, even if the results could be used to guide their own therapy (Yusuf et al., 2015). Nonwhites continue to have significantly lower rates of *BRCA1/2* testing uptake compared to their white counterparts (Levy et al., 2011). The unfortunate past history of ethical infringements on African Americans' rights in early experimentation for sickle cell disease, another genetic condition, may help explain why African Americans participate in genetic studies at an even lower rate than their participation in medical research generally (Royal et al., 2000). In addition, medical mistrust has been shown to impact how African American women engage in genetic counseling and testing (Sheppard et al., 2013). The

fact that there are very few African American genetics professionals may contribute further to the hesitancy of African Americans to undergo genetic testing (Halbert et al., 2006). Research is ongoing on how to approach African American and other minority populations about genetic testing (Bravo et al., 2014), but much remains to be learned about the factors responsible for the lack of involvement of minorities in genetic medicine to date.

It is also important to consider cultural health beliefs, although this topic has been understudied in the literature. One study found that immigrants were only one-third as likely to report a family history of cancer as nonimmigrants in a health survey of 5586 households in the United States, controlling for sociodemographic and cancer knowledge variables. Since a family history of cancer is an important factor affecting referral for genetic counseling and testing, this finding, which may relate to diminished knowledge of family history and/or to customs in the native country enforcing silence and secrecy about cancer diagnoses, may help explain the lower uptake of genetic services by individuals from ethnic minorities (Orom et al., 2008). It is not acculturation per se, but rather the strength of family beliefs and an openness of communication about cancer in the family that influence attitudes (Kenen et al., 2004; Barlow-Stewart et al., 2006; Werner-Lin, 2007). Cultural beliefs may also affect decision making about the uptake of risk-reduction options. Den Heijer et al. compared the attitudes about prophylactic mastectomy among physicians in France, Germany, the Netherlands, and the United Kingdom and found significant international variations, with the physicians from the Netherlands and the United Kingdom reporting more positive attitudes toward prophylactic mastectomy (Den Heijer et al., 2013).

25.3.4 Health Behaviors

While a 2002 Australian study suggested that after women were informed that they are *BRCA1/2* mutation carriers they were less likely to adhere to mammogram recommendations (Meiser et al., 2002), a French study found a significant uptake in screening practices after genetic testing (Maheu et al., 2012). A large population-based study utilizing a familial breast registry found that high-risk women with a greater perceived risk of developing breast cancer were more likely to have had a mammogram in the past 12 months and were more likely to practice breast self-exams once a month or more compared to participants who rated their risk as average or below average (Zhang et al., 2011).

Colon cancer screening compliance among the general population is lower than that for breast cancer, and studies have shown that populations at high risk for colon cancer are still not compliant with recommended screening (Lin et al., 2013). However, similar to the trend noted in breast cancer screening, there does appear to be an uptake in appropriate screening practices following genetic counseling for Lynch syndrome (Burton-Chase et al., 2013). One new randomized control trial demonstrated that a relatively brief but tailored intervention for individuals at high risk for colorectal cancer significantly increased screening adherence (Lowery et al., 2014). While this research is promising, much more research is needed regarding ways to mobilize subgroups of high-risk individuals who continue to be nonadherent to screening recommendations.

25.4 GETTING THE GENETIC TEST RESULTS: PERSONAL IMPACT AND PROFESSIONAL COMMUNICATION

The disclosure of a genetic test result is a moment of high drama. It is a moment of importance not only for the individual patient present and his or her spouse, but it also has implications for children, born and unborn, and grandchildren. There may also be implications, though less directly, for siblings, cousins, and other relatives. It may be the end of many years of wondering if one has a hereditary predisposition, or it may come as the response to a relatively recent education about hereditary illness, in some cases occasioned by a recent cancer diagnosis. Some results are more definitive than others. Learning that one is truly freed from the family curse by virtue of a true negative result can be a moment of blissful relief. Learning, alternatively, that one is a mutation carrier can answer for some people who have had cancer the question of why they became ill. However, it also opens up many other questions about the subsequent risks for the patient him/herself and questions about risk to offspring. For unaffected patients, learning that they are mutation carriers can be a fearsome moment, conjuring images of ill or deceased family members. These individuals often express disappointment that they could not be released from their cancer worry as they had imagined they would be if they had tested negative. The intensity of their immediate, emotional response often makes it difficult for patients to pay attention immediately to the recommendations for mutation carriers offered by the genetic counselor. For others, the result they receive is indeterminate. This can be both disappointing, as the answer is unclear, and relieving in that at least, for the moment, a mutation was not found. The latter response is somewhat irrational, since an indeterminate finding does not assume that a mutation is not present but only that one could not be found with present testing methods. In each case, emotions are aroused, and often patients recount that little was heard in the disclosure session beyond the result itself.

25.4.1 Testing Positive

The follow up of patients undergoing cancer genetic testing has shown that extreme adverse psychological outcomes

are rare (Hirschberg et al., 2015). Most patients who are found to be carriers of deleterious mutations respond with sadness but do not develop clinically significant symptoms of distress. A metaanalysis of 20 studies on the emotional consequences of predictive testing for *BRCA1/2* mutations conducted in 2009 found that while carriers' anxiety and cancer-specific distress increased immediately after receiving genetic testing results, the effect sizes were small and dissipated over time (Hamilton et al., 2009). Similar patterns were shown in individuals undergoing genetic testing for Lynch syndrome, with multiple prospective studies showing either no change in psychological distress or short-term increases in distress, with no long-term psychosocial effects of testing and with scores not reaching a clinical significance (Bleiker et al., 2013).

Qualitative studies that have explored the experience of teens and young adults who undergo predictive genetic testing have shown that the emotional outcomes of the disclosure often do not directly correlate with the testing results. Many of those who tested positive reported harms associated with the results, including witnessing distress in their parents, their own negative emotions, and concerns about the implications for future life, but they also indicated benefits, such as experiencing relief from uncertainty, being able to move forward with their lives, and strengthened friendships and family relationships (Duncan et al., 2008; Godino et al., 2015).

Disease status also affects psychological outcomes. The ability to correctly anticipate the level of one's emotional reaction to a forthcoming event is thought of as a predictor of good adjustment to that event. In a study of women tested for *BRCA1/2*, some of whom had had cancer and some who were unaffected (i.e., had not had cancer), the cancer patients found it significantly more difficult to anticipate their emotional reaction following disclosure (Dorval et al., 2000). Cancer patients tended to feel that learning that they were mutation carriers would not be as eventful for them as it turned out to be. An underestimation of the emotional impact of disclosure was associated with significantly increased distress levels 6 months later.

25.4.2 Testing Negative

In both HD and cancer genetic testing outcome studies, a subset of individuals with negative test results have been found to experience moderate levels of distress (Den Heijer et al., 2013; Bakos et al., 2008). These paradoxical reactions were largely unanticipated by researchers. Relief is prominent in the reactions of most individuals testing negative, although many (but not all) say they also feel guilty toward family members who have had the disease or who are mutation carriers (Macrae et al., 2013). In addition, many noncarriers describe that their experience of dealing with the burden of the disease does not end when they test negative, as they must still cope with the new diagnoses of other family members, and some continue to endorse moderate levels of cancer-specific worry (Macrae et al., 2013). More extreme negative reactions have occasionally been seen in individuals who are significantly distressed after learning that they are not mutation carriers due to regret for having lived so many years with the expectation of becoming ill and of having foregone opportunities for further education or other experiences because of that, now proved faulty, expectation (Huggins et al., 1992).

Another unexpected reaction of individuals testing negative for mutations in cancer genes or being told that their cancer risk is not sufficient to recommend genetic testing has been an unwillingness to accept the recommendations. Some find it difficult to give up the increased screening they had received before testing, when it was thought they could be mutation carriers. A study of female noncarriers from *BRCA* mutation-positive families found many underwent unnecessary mammography and ovarian ultrasounds compared to the general population (Dorval et al., 2011). Contrary to popular belief, anxiety has not been found to be associated with overscreening (Milhabet et al., 2013). However, in unaffected noncarriers of the *BRCA1/2* mutation, overscreening has been shown to be associated with feelings of vulnerability (Milhabet et al., 2013). Individuals testing negative for familial andenomatous polyposis have, in some cases, stated their reluctance to discontinue the early and more frequent colonoscopic screening recommended for those at high risk, stating insufficient trust in the genetic test result (Rhodes et al., 1991; Bleiker et al., 2003). Of note, many studies have found that subjective risk perception, not cancer risk based on family history, is associated with overscreening behaviors (Bleiker et al., 2013; Zhang et al., 2011).

25.4.3 Variants of Unknown Significance

There has been increased attention on individuals who receive a test result that shows them to have a variant of unknown significance (VUS). A VUS is an alteration in the normal sequence of a gene, the significance of which is currently unclear. As many as 40% of women who undergo *BRCA1/2* testing receive a test result of VUS (Tung et al., 2015). Miscomprehension and difficulty interpreting what a VUS means appear to be common (Hanoch et al., 2014; Richter et al., 2013) and may be higher in women with lower levels of education (Richter et al., 2013). This can have significant implications regarding distress, screening, and future psychological and physical health. Targeted educational interventions could help ensure an appropriate comprehension of testing results. Variants can, with time, be found to have clearer negative or positive significance. It is unclear to what extent patients understand this. Mechanisms for future contact with tested

patients who initially received VUS results often lack specification, so many patients may fail to gain important subsequent information about their risks. Research has also shown that VUS or inconclusive negative results are less likely to be communicated by women tested for breast/ovarian cancer genes to their sisters than more definitive positive or true negative results, likely due to difficulties describing and communicating the ambiguity of VUS results (Patenaude et al., 2006).

25.5 FAMILY COMMUNICATION

Genetic medicine is family medicine. Genetic professionals advise patients to consider carefully the ramifications of genetic information on other family members, and we continue to rely largely on tested probands to educate family members about hereditary risks. Both before and after genetic testing, there are reasons why open communication with family members about a family history and hereditary risk may be advisable. In order to provide the medical information genetics professionals require to calculate the risk that an individual carries a mutation, it is often necessary to talk to relatives about which family members were affected by which disease at what ages and with what outcomes. Women are typically the keepers of such family information (Green et al., 1997). Because of the many premature deaths in families with hereditary illness, there are often gaps in knowledge about the family history. Individuals seeking genetic counseling may have to talk to relatives with whom they ordinarily have little contact in order to complete the family history to a point where the counselor will have enough information to calculate risk.

Gathering family history information in order to create a family illness pedigree may require talking to relatives on both sides of the family, which sometimes includes uncomfortable contacts with long-lost relatives. Family traditions may include not talking very directly about illness, and doing so may open old emotional wounds. Some family members are completely estranged, making the completion of a full pedigree impossible. Because the hallmark of hereditary illness is the involvement of multiple generations of family members, family issues tend to surface more readily with regard to genetic testing than in other kinds of medical endeavors. Since cancer genetic testing is now much more frequently conducted on individuals with a less striking family cancer history, gathering accurate information about cancers that did occur is even more crucial.

The relevance of one person's genetic information to other family members can set up difficult alliances or the tendency to avoid relatives with whom one is not comfortable discussing highly personal information, such as a genetic test result. The inaccuracy of family information can affect the genetic professional's ability to correctly analyze the likelihood that there is a hereditary predisposition to disease in a particular family. The documentation of the age of onset of a disease is particularly important in such determinations, since, in many cases, an early age of onset is a hallmark of hereditary illness. Also, knowledge of exactly which cancers a family member had may be critical, as cancer syndromes link certain cancers to hereditary mutations. There is more confusion about some cancers than others. Knowledge about breast cancer within a family, for example, tends to be more accurate than colon cancer, which may be confused with other gastrointestinal cancers (Glanz et al., 1999; Schneider et al., 2004). These reports suggest the need to verify diagnoses from medical records of the family members, if possible, as part of the genetic counseling process.

Although most inaccuracies are due to misconceptions or unintentional errors, there are rare occasions when family embarrassment about illness or other factors lead to a deliberate falsification of the family history (Evans et al., 1996). Another risk of gathering family history data is the risk of uncovering misassigned paternity; that is, that the putative father of a child is not, in fact, the biological father. Patients should be helped to understand that the experience of searching for family history data is, in some cases, more emotional and/or upsetting than anticipated.

Once genetic test results are available and disclosed to the patient, considerations about the implications of the result for other family members come into play. Genetics professionals recommend quite strongly that results be shared with family members, especially when screening recommendations likely would be influenced by the genetic information. For example, if a man in his 50s is found to be a mutation carrier for HNPCC, it would be important for his children to know that they should initiate colonoscopy screenings at age 25, not at age 50, as recommended for the general population. Similarly, if the man is negative, his children could use that information to decide that they do not require colonoscopy until they are 50 years of age.

In most cases, information about genetic testing is readily shared, but this is not universal. There are rifts in many families that prevent parents, children, and siblings from communicating about difficult subjects such as health risks. Some do not talk at all. Divorce, disgruntlement, and embarrassment can all inhibit or prevent an open discussion of such mutually relevant medical information. Other family crises, including disease-related deaths, may impede an immediate disclosure and discussion of genetic test results. It can be difficult to decide how and when to inform relatives. Since the testing of children for adult-onset disorders is generally discouraged, except when the results would influence an uptake of a preventive option that would begin in childhood, parents often wonder when or how they should share their test results with their young children. Early research in this area,

however, has shown that it is the family's style of communicating generally that typically governs whether and when children are told (Tercyak et al., 2002). In *BRCA1/2* families, two-thirds of minor children are told their mother's test result within a few months after disclosure (Tercyak et al., 2013). Negative or uninformative results are more likely to be shared with minor children than positive results, and teenage children are more likely to be told than younger children. Telling seems to relieve most parents who feel that they have discharged an important parental duty (Patenaude et al., 2013b). There are many suggestions that parents would value a decision aid in preparing to talk to their children and answer their questions (Patenaude et al., 2013b).

Some family members do not wish to know about their own or others' genetic test status. This can raise problems, especially when that family member is a twin or an obligate carrier of a mutation identified in another relative. Some family members are fearful that telling a relative could be quite upsetting and bring back unpleasant memories or arouse feelings of guilt for having passed on a deleterious mutation. Males continue to lag in the receipt of results of others in the family (Patenaude et al., 2006; Montgomery et al., 2013; Vadaparampil et al., 2012). In a study designed to assess family communication and test uptake among relatives of individuals of lower socioeconomic status and greater diversity than those typically studied, it was shown that tested individuals of African American and Asian/Pacific Islander descent were less likely to disclose genetic test results to their relatives, and the relatives were less likely to be tested than were relatives of Caucasian, higher socioeconomic status probands (Fehniger et al., 2013). The communication of test results in families affected by Lynch syndrome shows lower rates of communicating than for *BRCA1/2* test results (Menko et al., 2013).

Providers need help in advising patients about the communication of results to family members. As the genetic component of many diseases is recognized, the complexity of such communication will increase. Both patients and providers will find it valuable to utilize aids to communicate complicated genetic information to relatives, but this will not replace the need for sensitive patient−provider communication and awareness of the complex psychological responses that the dissemination of genetic information within a family can evoke (Rupert et al., 2013).

25.6 FUTURE CHALLENGES: COMPLEXITY AND DIVERSITY

Genetic testing and genomic sequencing promise to add many layers of complexity for both patients and providers going forward. The number of genes we can test for are likely to increase faster than our ability to clearly interpret the results. While to date we have tested largely for genes where we understand the penetrance and the range of diseases caused by the deleterious mutation, this is much less likely to be the case in future testing. This raises issues of enormous complexity for providers trying to explain the potential value of testing and to interpret results for perplexed patients. In turn, it becomes much more difficult for patients, to whom we still entrust the task of informing relatives, to faithfully convey the meaning of a test result they (and possibly their providers) do not fully understand. The need for educational aids will grow inversely; the less we know clearly, the more patients will need help explaining what a genetic test or genomic sequence tells us about our health risks.

Technology is likely to play a larger role in genetic testing and counseling. Telephone counseling and result delivery are increasing (Kinney et al., 2014; Sie et al., 2014). Internet access is fast replacing face-to-face contact for pedigree gathering and psychoeducational intervention. It will be important that we do not lose the ability to listen to the patient and to convey the necessary sense of empathy and long-term commitment to supporting and educating patients in the process. There will always be a need for a sympathetic and respectful clinician who can help patients make links between their family history, their genetic risk, and the possibilities of reducing adverse outcomes of hereditary predisposition (McClellan et al., 2013).

Genetic medicine holds great promise for identifying individuals at an increased disease risk and for treating patients with greater efficacy and fewer side effects through pharmacogenetics. Patient education about genetics will necessitate having well-informed, well-trained, culturally competent professionals to address the different genetic counseling, medical, and psychological questions that will arise (Weitzel et al., 2011). It is a challenge for professionals to educate themselves about this rapidly changing field. In discussing and delineating the risks and benefits of genetic testing, it will also be important to discuss the psychological factors governing the patient and family response to genetic information. It is only when patients feel empowered and not threatened by such information that they can make optimal use of the advances in genetic medicine.

REFERENCES

Bakos, A.D., Hutson, S.P., Loud, J.T., Peters, J.A., Giusti, R.M., Greene, M.H., 2008. BRCA mutation-negative women from hereditary breast and ovarian cancer families: a qualitative study of the BRCA-negative experience. Health Expect. 11 (3), 220−231.

Barlow-Stewart, K., Yeo, S.S., Meiser, B., Goldstein, D., Tucker, K., Eisenbruch, M., 2006. Toward cultural competence in cancer genetic

counseling and genetics education: lessons learned from Chinese-Australians. Genet. Med. 8 (1), 24–32.

Bleiker, E., Menko, F., Taal, B., et al., 2003. Experience of discharge from colonoscopy of mutation negative HNPCC family members. J. Med. Genet. 40 (5), e55 e.

Bleiker, E.M., Esplen, M.J., Meiser, B., Petersen, H.V., Patenaude, A.F., 2013. 100 years Lynch syndrome: what have we learned about psychosocial issues? Fam. Cancer 12 (2), 325–339.

Bravo, C., O'Donoghue, C., Kaplan, C.P., Luce, J., Ozanne, E., 2014. Can mHealth improve risk assessment in underserved populations? Acceptability of a breast health questionnaire App in ethnically diverse, older, low-income women. J. Health Dispar. Res. Pract. 7 (4).

Bredart, A., Kop, J.L., Depauw, A., et al., 2013. Short-term psychological impact of the BRCA1/2 test result in women with breast cancer according to their perceived probability of genetic predisposition to cancer. Br. J. Cancer 108 (5), 1012–1020.

Burton-Chase, A.M., Hovick, S.R., Peterson, S.K., et al., 2013. Changes in screening behaviors and attitudes toward screening from pre-test genetic counseling to post-disclosure in Lynch syndrome families. Clin. Genet. 83 (3), 215–220.

Butow, P., Meiser, B., Price, M., et al., 2005. Psychological morbidity in women at increased risk of developing breast cancer: a controlled study. Psychooncology 14 (3), 196–203.

Cella, D., Hughes, C., Peterman, A., et al., 2002. A brief assessment of concerns associated with genetic testing for cancer: the Multidimensional Impact of Cancer Risk Assessment (MICRA) questionnaire. Health Psychol. 21 (6), 564.

Cherry, M.C., Ropka, M., Lyle, M.J., Napolitano, L., Daly, M.B., 2013. Understanding the needs of women considering risk reducing Salpingo-Oophorectomy. Cancer Nurs. 36 (3), E33.

Claes, E., Evers-Kiebooms, G., Decruyenaere, M., et al., 2005. Surveillance behavior and prophylactic surgery after predictive testing for hereditary breast/ovarian cancer. Behav. Med. 31 (3), 93–106.

Collins, F.S., Green, E.D., Guttmacher, A.E., Guyer, M.S., 2003. A vision for the future of genomics research. Nature 422 (6934), 835–847.

David, S.P., Johnson, S.G., Berger, A.C., et al., 2015. Making personalized health care even more personalized: insights from activities of the IOM genomics roundtable. Ann. Fam. Med. 13 (4), 373–380.

Davis, J.L., Buchanan, K.L., Green, B.L., 2013. Racial/ethnic differences in cancer prevention beliefs: applying the health belief model framework. Am. J. Health Promot. 27 (6), 384–389.

Den Heijer, M., van Asperen, C.J., Harris, H., et al., 2013. International variation in physicians' attitudes towards prophylactic mastectomy—comparison between France, Germany, The Netherlands and the United Kingdom. Eur. J. Cancer 49 (13), 2798–2805.

Domanska, K., Nilbert, M., Soller, M., Silfverberg, B., Carlsson, C., 2007. Discrepancies between estimated and perceived risk of cancer among individuals with hereditary nonpolyposis colorectal cancer. Genet. Test. 11 (2), 183–186.

Dorval, M., Patenaude, A.F., Schneider, K.A., et al., 2000. Anticipated versus actual emotional reactions to disclosure of results of genetic tests for cancer susceptibility: findings from p53 and BRCA1 testing programs. J. Clin. Oncol. 18 (10), 2135–2142.

Dorval, M., Nogues, C., Berthet, P., et al., 2011. Breast and ovarian cancer screening of non-carriers from BRCA1/2 mutation-positive families: 2-year follow-up of cohorts from France and Quebec. Eur. J. Hum. Genet. 19 (5), 494–499.

Douma, K.F., Aaronson, N.K., Vasen, H.F., et al., 2010. Psychological distress and use of psychosocial support in familial adenomatous polyposis. Psychooncology 19 (3), 289–298.

Duncan, R.E., Gillam, L., Savulescu, J., Williamson, R., Rogers, J.G., Delatycki, M.B., 2008. "You're one of us now": young people describe their experiences of predictive genetic testing for Huntington disease (HD) and familial adenomatous polyposis (FAP). Am. J. Med. Genet. Part C Seminars Med. Genet. 47–55. Wiley Online Library.

Easton, D.F., Pharoah, P.D., Antoniou, A.C., et al., 2015. Gene-panel sequencing and the prediction of breast-cancer risk. N. Engl. J. Med. 372 (23), 2243–2257.

Edwards, K.T., Huang, C.J., 2014. Bridging the consumer-medical divide: how to regulate direct-to-consumer genetic testing. Hastings Cent. Rep. 44 (3), 17–19.

Evans, D.G.R., Kerr, B., Cade, D., Hoare, E., Hopwood, P., 1996. Fictitious breast cancer family history. Lancet 348 (9033), 1034.

Evans, D.G., Barwell, J., Eccles, D.M., et al., 2014. The Angelina Jolie effect: how high celebrity profile can have a major impact on provision of cancer related services. Breast Cancer Res. 16 (5), 442.

Feero, W.G., Guttmacher, A.E., Collins, F.S., 2008. The genome gets personal—almost. JAMA 299 (11), 1351–1352.

Fehniger, J., Lin, F., Beattie, M.S., Joseph, G., Kaplan, C., 2013. Family communication of BRCA1/2 results and family uptake of BRCA1/2 testing in a diverse population of BRCA1/2 carriers. J. Genet. Couns. 22 (5), 603–612.

Geransar, R., Einsiedel, E., 2008. Evaluating online direct-to-consumer marketing of genetic tests: informed choices or buyers beware? Genet. Test. 12 (1), 13–23.

Glanz, K., Grove, J., Le Marchand, L., Gotay, C., 1999. Underreporting of family history of colon cancer: correlates and implications. Cancer Epidemiol. Biomarkers Prev. 8 (7), 635–639.

Godino, L., Turchetti, D., Jackson, L., Hennessy, C., Skirton, H., 2015. Impact of presymptomatic genetic testing on young adults: a systematic review. Eur. J. Hum. Genet. 24 (4), 496–503.

Graves, K.D., Vegella, P., Poggi, E.A., et al., 2012. Long-term psychosocial outcomes of BRCA1/BRCA2 testing: differences across affected status and risk-reducing surgery choice. Cancer Epidemiol. Biomarkers Prev. 21 (3), 445–455.

Green, J., Richards, M., Murton, F., Statham, H., Hallowell, N., 1997. Family communication and genetic counseling: the case of hereditary breast and ovarian cancer. J. Genet. Couns. 6 (1), 45–60.

Hadley, D.W., Jenkins, J.F., Steinberg, S.M., et al., 2008. Perceptions of cancer risks and predictors of colon and endometrial cancer screening in women undergoing genetic testing for Lynch syndrome. J. Clin. Oncol. 26 (6), 948–954.

Halbert, C.H., Kessler, L., Stopfer, J.E., Domchek, S., Wileyto, E.P., 2006. Low rates of acceptance of BRCA1 and BRCA2 test results among African American women at increased risk for hereditary breast-ovarian cancer. Genet. Med. 8 (9), 576–582.

Hallowell, N., Statham, H., Murton, F., 1998. Women's understanding of their risk of developing breast/ovarian cancer before and after genetic counseling. J. Genet. Couns. 7 (4), 345–364.

Hamilton, J.G., Lobel, M., Moyer, A., 2009. Emotional distress following genetic testing for hereditary breast and ovarian cancer: a meta-analytic review. Health Psychol. 28 (4), 510.

Hanoch, Y., Miron-Shatz, T., Rolison, J.J., Ozanne, E., 2014. Understanding of BRCA1/2 genetic tests results: the importance of objective and subjective numeracy. Psychooncology 23 (10), 1142–1148.

Hirschberg, A.M., Chan-Smutko, G., Pirl, W.F., 2015. Psychiatric implications of cancer genetic testing. Cancer 121 (3), 341–360.

Hoskins, L.M., Roy, K.M., Greene, M.H., 2012. Toward a new understanding of risk perception among young female BRCA1/2 "previvors". Fam. Syst. Health 30 (1), 32–46.

Hudson, K.L., Holohan, M., Collins, F.S., 2008. Keeping pace with the times—the genetic information nondiscrimination act of 2008. N. Engl. J. Med. 358 (25), 2661–2663.

Huggins, M., Bloch, M., Wiggins, S., et al., 1992. Predictive testing for Huntington disease in Canada: adverse effects and unexpected results in those receiving a decreased risk. Am. J. Med. Genet. 42 (4), 508–515.

Iglehart, J.D., Miron, A., Rimer, B.K., Winer, E.P., Berry, D., Shildkraut, M., 1998. Overestimation of hereditary breast cancer risk. Ann. Surg. 228 (3), 375.

Jolie, A., 2013. My medical choice. N.Y. Times 14 (05).

Juthe, R.H., Zaharchuk, A., Wang, C., 2015. Celebrity disclosures and information seeking: the case of Angelina Jolie. Gen. Med. 17 (7), 545–553.

Kash, K.M., Ortega-Verdejo, K., Dabney, M.K., Holland, J.C., Miller, D.G., Osborne, M.P., 2000. Psychosocial aspects of cancer genetics: women at high risk for breast and ovarian cancer. Semin. Surg. Oncol. 333–338. Wiley Online Library.

Kenen, R., Arden-Jones, A., Eeles, R., 2004. We are talking, but are they listening? Communication patterns in families with a history of breast/ovarian cancer (HBOC). Psychooncology 13 (5), 335–345.

Kinney, A.Y., Boonyasiriwat, W., Walters, S.T., et al., 2014. Telehealth personalized cancer risk communication to motivate colonoscopy in relatives of patients with colorectal cancer: the family CARE Randomized controlled trial. J. Clin. Oncol. 32 (7), 654–662.

Lammens, C.R., Aaronson, N.K., Wagner, A., et al., 2010. Genetic testing in Li-Fraumeni syndrome: uptake and psychosocial consequences. J. Clin. Oncol. 28 (18), 3008–3014.

Levy, D.E., Byfield, S.D., Comstock, C.B., et al., 2011. Underutilization of BRCA1/2 testing to guide breast cancer treatment: black and Hispanic women particularly at risk. Genet. Med. 13 (4), 349–355.

Lin, O.S., Gluck, M., Nguyen, M., Koch, J., Kozarek, R.A., 2013. Screening patterns in patients with a family history of colorectal cancer often do not adhere to national guidelines. Dig. Dis. Sci. 58 (7), 1841–1848.

Lowery, J.T., Horick, N., Kinney, A.Y., et al., 2014. A randomized trial to increase colonoscopy screening in members of high-risk families in the colorectal cancer family registry and cancer genetics network. Cancer Epidemiol. Biomarkers Prev. 23 (4), 601–610.

Lu, K.H., 2015. New Strategies for Ovarian Cancer Prevention in High Risk Women: Dana-Farber Cancer Institute.

Macrae, L., de Souza, A.N., Loiselle, C.G., Wong, N., 2013. Experience of BRCA1/2 mutation-negative young women from families with hereditary breast and ovarian cancer: a qualitative study. Hered. Cancer Clin. Pract. 11 (1), 14.

Maheu, C., Apostolidis, T., Petri-Cal, A., et al., 2012. French women's breast self-examination practices with time after undergoing BRCA1/2 genetic testing. Fam. Cancer 11 (2), 269–278.

Manchanda, R., Loggenberg, K., Sanderson, S., et al., 2015. Population testing for cancer predisposing BRCA1/BRCA2 mutations in the Ashkenazi-Jewish community: a randomized controlled trial. J. Natl. Cancer Inst. 107 (1), 379.

McClellan, K.A., Kleiderman, E., Black, L., et al., 2013. Exploring resources for intrafamilial communication of cancer genetic risk: we still need to talk. Eur. J. Hum. Genet. 21 (9), 903–910.

McInerney-Leo, A., Hadley, D., Kase, R.G., Giambarresi, T.R., Struewing, J.P., Biesecker, B.B., 2006. BRCA1/2 testing in hereditary breast and ovarian cancer families III: risk perception and screening. Am. J. Med. Genet. Part A 140 (20), 2198–2206.

Meiser, B., Butow, P., Friedlander, M., et al., 2002. Psychological impact of genetic testing in women from high-risk breast cancer families. Eur. J. Cancer 38 (15), 2025–2031.

Menko, F.H., Aalfs, C.M., Henneman, L., et al., 2013. Informing family members of individuals with Lynch syndrome: a guideline for clinical geneticists. Fam. Cancer 12 (2), 319–324.

Milhabet, I., Duprez, C., Krzeminski, A., Christophe, V., 2013. Cancer risk comparative perception and overscreening behaviours of non-carriers from BRCA1/2 families. Eur. J. Cancer Care 22 (4), 540–548.

Miller, S.M., Roussi, P., Caputo, G.C., Kruus, L., 1995. Patterns of children's coping with an aversive dental treatment. Health Psychol. 14 (3), 236.

Montgomery, S.V., Barsevick, A.M., Egleston, B.L., et al., 2013. Preparing individuals to communicate genetic test results to their relatives: report of a randomized control trial. Fam. Cancer 12 (3), 537–546.

Murakami, Y., Okamura, H., Sugano, K., et al., 2004. Psychologic distress after disclosure of genetic test results regarding hereditary nonpolyposis colorectal carcinoma. Cancer 101 (2), 395–403.

Myers, M.F., Chang, M.-H., Jorgensen, C., et al., 2006. Genetic testing for susceptibility to breast and ovarian cancer: evaluating the impact of a direct-to-consumer marketing campaign on physicians' knowledge and practices. Genet. Med. 8 (6), 361–370.

Nelson, H.D., Pappas, M., Zakher, B., Mitchell, J.P., Okinaka-Hu, L., Fu, R., 2014. Risk assessment, genetic counseling, and genetic testing for BRCA-related cancer in women: a systematic review to update the US preventive services task force recommendation. Ann. Intern. Med. 160 (4), 255–266.

Noar, S.M., Althouse, B.M., Ayers, J.W., Francis, D.B., Ribisl, K.M., 2015. Cancer information seeking in the digital age: effects of Angelina Jolie's prophylactic mastectomy announcement. Med. Decis. Making 35 (1), 16–21.

Offit, K., Geffen, D., 1998. Clinical cancer genetics: risk counselling and management. Ann. Oncol. 9 (10), 1142.

Orom, H., Cote, M.L., Gonzalez, H.M., Underwood 3rd, W., Schwartz, A.G., 2008. Family history of cancer: is it an accurate indicator of cancer risk in the immigrant population? Cancer 112 (2), 399–406.

Orom, H., Kiviniemi, M.T., Underwood, W., Ross, L., Shavers, V.L., 2010. Perceived cancer risk: why is it lower among nonwhites than whites? Cancer Epidemiol. Biomarkers Prev. 19 (3), 746–754.

Patenaude, A.F., Dorval, M., DiGianni, L.S., Schneider, K.A., Chittenden, A., Garber, J.E., 2006. Sharing BRCA1/2 test results with first-degree relatives: factors predicting who women tell. J. Clin. Oncol. 24 (4), 700–706.

Patenaude, A.F., Tung, N., Ryan, P.D., et al., 2013. Young adult daughters of BRCA1/2 positive mothers: what do they know about hereditary cancer and how much do they worry? Psychooncology 22 (9), 2024–2031.

Patenaude, A.F., DeMarco, T.A., Peshkin, B.N., et al., 2013. Talking to children about maternal BRCA1/2 genetic test results: a qualitative study of parental perceptions and advice. J. Genet. Couns. 22 (3), 303−314.

Phimister, E.G., Feero, W.G., Guttmacher, A.E., 2012. Realizing genomic medicine. N. Engl. J. Med. 366 (8), 757−759.

Rhodes, M., Chapman, P., Burn, J., Gunn, A., 1991. Role of a regional register for familial adenomatous polyposis: experience in the Northern Region. Br. J. Surg. 78 (4), 451−452.

Richter, S., Haroun, I., Graham, T., Eisen, A., Kiss, A., Warner, E., 2013. Variants of unknown significance in BRCA testing: impact on risk perception, worry, prevention and counseling. Ann. Oncol. 24 (Suppl. 8), viii69−viii74.

Robson, M.E., Bradbury, A.R., Arun, B., et al., 2015. American Society of clinical oncology Policy Statement update: genetic and genomic testing for Cancer susceptibility. J. Clin. Oncol. 33 (31), 3660−3667.

Roussi, P., Miller, S.M., 2014. Monitoring style of coping with cancer related threats: a review of the literature. J. Behav. Med. 37 (5), 931−954.

Roussi, P., Sherman, K.A., Miller, S.M., et al., 2011. Identification of cognitive profiles among women considering BRCA1/2 testing through the utilisation of cluster analytic techniques. Psychol. Health 26 (10), 1327−1343.

Royal, C., Baffoe-Bonnie, A., Kittles, R., et al., 2000. Recruitment experience in the first phase of the African American hereditary Prostate Cancer (AAHPC) study. Ann. Epidemiol. 10 (8), S68−S77.

Rupert, D.J., Squiers, L.B., Renaud, J.M., et al., 2013. Communicating risk of hereditary breast and ovarian cancer with an interactive decision support tool. Patient Educ. Couns. 92 (2), 188−196.

Schneider, K.A., DiGianni, L.M., Patenaude, A.F., et al., 2004. Accuracy of cancer family histories: comparison of two breast cancer syndromes. Genet. Test. 8 (3), 222−228.

Schwartz, M.D., Peshkin, B.N., Hughes, C., Main, D., Isaacs, C., Lerman, C., 2002. Impact of BRCA1/BRCA2 mutation testing on psychologic distress in a clinic-based sample. J. Clin. Oncol. 20 (2), 514−520.

Schwartz, M.D., Lerman, C., Brogan, B., et al., 2004. Impact of BRCA1/BRCA2 counseling and testing on newly diagnosed breast cancer patients. J. Clin. Oncol. 22 (10), 1823−1829.

Sheppard, V.B., Mays, D., LaVeist, T., Tercyak, K.P., 2013. Medical mistrust and self-efficacy influence Black women's level of engagement in BRCA1/2 genetic counseling and testing. J. Natl. Med. Assoc. 105 (1), 17.

Shiloh, S., Koehly, L., Jenkins, J., Martin, J., Hadley, D., 2008. Monitoring coping style moderates emotional reactions to genetic testing for hereditary nonpolyposis colorectal cancer: a longitudinal study. Psychooncology 17 (8), 746−755.

Sie, A.S., van Zelst-Stams, W.A., Spruijt, L., et al., 2014. More breast cancer patients prefer BRCA-mutation testing without prior face-to-face genetic counseling. Fam. Cancer 13 (2), 143−151.

Slavin, T.P., Niell-Swiller, M., Solomon, I., et al., 2015. Clinical Application of multigene panels: challenges of next-generation counseling and Cancer risk management. Front. Oncol. 5, 208.

Tercyak, K.P., Peshkin, B.N., DeMarco, T.A., Brogan, B.M., Lerman, C., 2002. Parent−child factors and their effect on communicating BRCA1/2 test results to children. Patient Educ. Couns. 47 (2), 145−153.

Tercyak, K.P., Peshkin, B.N., Brogan, B.M., et al., 2007. Quality of life after contralateral prophylactic mastectomy in newly diagnosed high-risk breast cancer patients who underwent BRCA1/2 gene testing. J. Clin. Oncol. 25 (3), 285−291.

Tercyak, K.P., Mays, D., DeMarco, T.A., et al., 2013. Decisional outcomes of maternal disclosure of BRCA1/2 genetic test results to children. Cancer Epidemiol. Biomarkers Prev. 22 (7), 1260−1266.

Tung, N., Battelli, C., Allen, B., et al., 2015. Frequency of mutations in individuals with breast cancer referred for BRCA1 and BRCA2 testing using next-generation sequencing with a 25-gene panel. Cancer 121 (1), 25−33.

Turner-Cobb, J.M., Bloor, L.E., Whittemore, A.S., West, D., Spiegel, D., 2006. Disengagement and social support moderate distress among women with a family history of breast cancer. Breast J. 12 (1), 7−15.

Vadaparampil, S.T., Malo, T., de la Cruz, C., Christie, J., 2012. Do breast Cancer patients tested in the oncology care setting share BRCA mutation results with family members and health care providers? J. Cancer Epidemiol. 2012, 498062.

van Oostrom, I., Meijers-Heijboer, H., Duivenvoorden, H.J., et al., 2006. Experience of parental cancer in childhood is a risk factor for psychological distress during genetic cancer susceptibility testing. Ann. Oncol. 17 (7), 1090−1095.

van Oostrom, I., Meijers-Heijboer, H., Duivenvoorden, H.J., et al., 2007. Prognostic factors for hereditary cancer distress six months after BRCA1/2 or HNPCC genetic susceptibility testing. Eur. J. Cancer 43 (1), 71−77.

Vernon, S.W., Gritz, E.R., Peterson, S.K., et al., 1997. Correlates of psychologic distress in colorectal cancer patients undergoing genetic testing for hereditary colon cancer. Health Psychol. 16 (1), 73.

Vos, J., Stiggelbout, A.M., Oosterwijk, J., et al., 2011. A counselee-oriented perspective on risk communication in genetic counseling: explaining the inaccuracy of the counselees' risk perception shortly after BRCA1/2 test result disclosure. Genet. Med. 13 (9), 800−811.

Vos, J., Gomez-Garcia, E., Oosterwijk, J.C., et al., 2012. Opening the psychological black box in genetic counseling. The psychological impact of DNA testing is predicted by the counselees' perception, the medical impact by the pathogenic or uninformative BRCA1/2-result. Psychooncology 21 (1), 29−42.

Vos, J., Oosterwijk, J.C., Gomez-Garcia, E., et al., 2012. Exploring the short-term impact of DNA-testing in breast cancer patients: the counselees' perception matters, but the actual BRCA1/2 result does not. Patient Educ. Couns. 86 (2), 239−251.

Watson, M., Kash, K., Homewood, J., Ebbs, S., Murday, V., Eeles, R., 2005. Does genetic counseling have any impact on management of breast cancer risk? Genet. Test. 9 (2), 167−174.

Weitzel, J.N., McCaffrey, S.M., Nedelcu, R., MacDonald, D.J., Blazer, K.R., Cullinane, C.A., 2003. Effect of genetic cancer risk assessment on surgical decisions at breast cancer diagnosis. Arch. Surg. 138 (12), 1323−1328.

Weitzel, J.N., Blazer, K.R., MacDonald, D.J., Culver, J.O., Offit, K., 2011. Genetics, genomics, and cancer risk assessment: state of the Art and future directions in the Era of personalized medicine. CA Cancer J. Clin. 61 (5), 327−359.

Werner-Lin, A.V., 2007. Danger zones: risk perceptions of young women from families with hereditary breast and ovarian cancer. Fam. Process 46 (3), 335−349.

Wevers, M.R., Hahn, D.E., Verhoef, S., et al., 2012. Breast cancer genetic counseling after diagnosis but before treatment: a pilot study on

treatment consequences and psychological impact. Patient Educ. Couns. 89 (1), 89−95.

Williams-Jones, B., 2003. Where there'sa web, there'sa way: commercial genetic testing and the internet. Public Health Genomics 6 (1), 46−57.

Yusuf, R.A., Rogith, D., Hovick, S.R., et al., 2015. Attitudes toward molecular testing for personalized cancer therapy. Cancer 121 (2), 243−250.

Zakowski, S.G., Valdimarsdottir, H.B., Bovbjerg, D.H., et al., 1997. Predictors of intrusive thoughts and avoidance in women with family histories of breast cancer. Ann. Behav. Med. 19 (4), 362−369.

Zhang, L.R., Chiarelli, A.M., Glendon, G., et al., 2011. Influence of perceived breast cancer risk on screening behaviors of female relatives from the Ontario site of the breast cancer family registry. Eur. J. Cancer Prev. 20 (4), 255.

Zilliacus, E., Meiser, B., Gleeson, M., et al., 2012. Are we being overly cautious? A qualitative inquiry into the experiences and perceptions of treatment-focused germline BRCA genetic testing amongst women recently diagnosed with breast cancer. Support. Care Cancer 20 (11), 2949−2958.

Chapter 26

Genomic Medicine in Developing Countries and Resource-Limited Environments

T. Katsila[1,a], K. Mitropoulos[2,a], Z. Mohamed[3], D.A. Forero[4], P. Laissue[5], A. Wonkam[6], C. Lopez-Correa[7], W. Chantratita[8], A. Llerena[9], B.R. Ali[10] and G.P. Patrinos[1,10]

[1]*University of Patras School of Health Sciences, Patras, Greece;* [2]*The Golden Helix Foundation, London, United Kingdom;* [3]*University of Malaya, Kuala Lumpur, Malaysia;* [4]*Universidad Antonio Nariño, Bogotá, Colombia;* [5]*Del Rosario University, Bogotá, Colombia;* [6]*University of Cape Town, Cape Town, South Africa;* [7]*Genome Quebec, Montreal, QC, Canada;* [8]*Mahidol University, Bangkok, Thailand;* [9]*Extremadura University Hospital and Medical School, Badajoz, Spain;* [10]*United Arab Emirates University, Al-Ain, United Arab Emirates*

26.1 INTRODUCTION

The advent of genomics and related technologies has revolutionized mainstream medical practices (McCarthy et al., 2013). The advent of genome-wide studies has catalyzed the translation of genomic findings into health care (Mardis, 2011). New technologies, particularly next-generation sequencing (NGS) approaches, are being adopted by diagnostic laboratories and hospitals in the United States and Western Europe. In terms of regulation, several guidelines from the United States Food and Drug Administration (http://www.fda.gov) and the European Medicines Agency (http://www.ema.europa.eu) are being announced regarding the translation of genomic medicine into the clinic.

However, genomic medicine is implemented at a different pace when developing and resource-limited countries are considered. In these countries, significant barriers exist, which often relate to limited resources and a lack of technology and knowledge transfer. As such, the potential of genomic medicine is often hardly understood by biomedical scientists and healthcare professionals. Considering that approximately 85% of the world's population lives in developing/resource-limited countries, access to genomic medicine becomes fundamental. Today, there have been some examples from the successful implementation of genomic medicine in developing countries in Europe and Asia that rely on several related and intersected disciplines (population genomics, pharmacogenomics (PGx), informatics, and public health genomics). We feel that the examples described in the following relate to the previously mentioned disciplines and can serve as model cases toward the implementation of genomic medicine in resource-limited environments.

26.1.1 Euro-PGx Project: A European-Wide Pharmacogenomics Map

In Europe, PGx is implemented in the various health systems at a rather heterogeneous pace. This is due to the lack of harmonization of the national guidelines within Europe and, most importantly, differences in resource availability (Mitropoulos et al., 2011). Furthermore, taking into account that the pharmacogenomic biomarker allele frequencies in various European populations are hardly known, it becomes challenging to define the actionable pharmacogenomic biomarkers on which drug—dose recommendations will be set. The Euro-PGx project (http://www.goldenhelix.org/index.php/research/pharmacogenomics-in-europe) focuses on the determination of the varying pharmacogenomic biomarker allele frequencies in a large number of mostly developing European countries to produce drug—dose recommendations. Preliminary findings from a large-

[a] These authors contributed equally to this work.

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00026-2
Copyright © 2017 Elsevier Ltd. All rights reserved.

scale genotyping effort using the Drug-Metabolizing Enzymes and Transporters Plus microarray (Affymetrix, Santa Clara, CA, USA) suggest that several pharmacogenomic biomarker allele frequencies vary significantly, despite the strong Caucasian genetic component of the vast majority of the European populations. Data access to the scientific community is anticipated through the FINDbase database (http://www.findbase.org) via the microattribution approach (Giardine et al., 2011). The Euro-PGx project also facilitates the organization of PGx educational activities in various European countries (as of May 2016, 16 different events have been organized in 10 European, mostly developing, countries). These educational events, also known as the Golden Helix PGx Days (http://pharmacogenomicsdays.goldenhelix.org), aim to educate healthcare professionals and increase PGx awareness.

26.1.2 Implementation of Pharmacogenomics in Clinical Settings

In Southeast Asia, particularly Taiwan, the associations of the HLA-B*1502 allele with Stevens-Johnson Syndrome/Toxic Epidermal Necrolysis (SJS/10) upon carbamazepine administration (Chung et al., 2004) as well as that of the HLA-B*5801 allele with allopurinol-induced severe cutaneous adverse reactions (Hung et al., 2005) support the value of PGx in tailor-made therapeutics. A large clinical study that confirmed the benefit of HLA-B*1502 screening to prospectively identify subjects at genetic risk for the previously mentioned condition (Chen et al., 2011) led the Taiwanese government to begin reimbursing the screening costs in 2010.

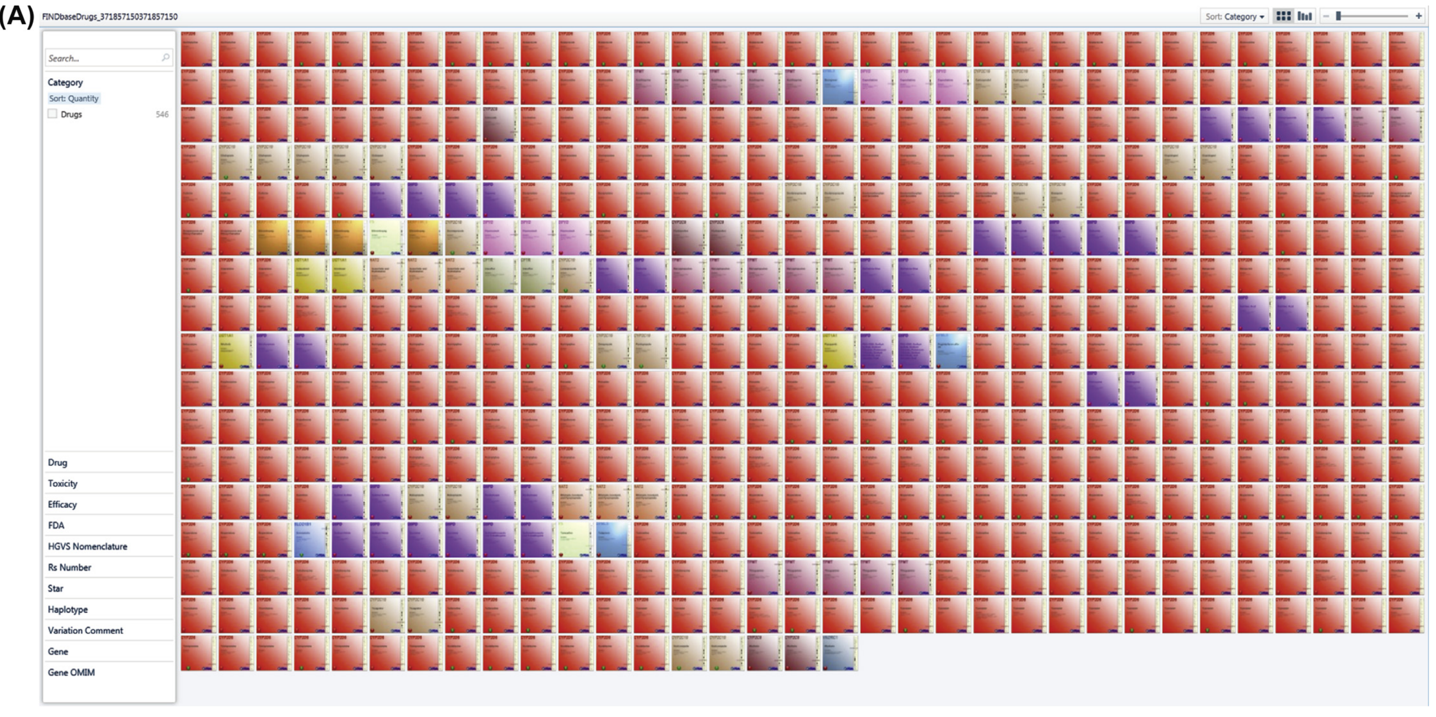
Today, several PGx biomarkers have been found to be correlated with interindividual drug efficacy and/or toxicity. Nevertheless, health professionals often lack genomics education, so it is urgent to make PGx knowledge readily available in a user-friendly format. The DruGeVar database (http://drugevar.genomicmedicinealliance.org) (Dalabira et al., 2014) was developed to serve as an online knowledge portal for clinical PGx with the aim of triangulating drugs−genes−PGx biomarkers (those approved by regulatory agencies) (Fig. 26.1).

In Southeast Asia, a pharmacogenomic card has been proposed to record patients' pharmacogenomic biomarkers. Similarly, the Ramathibodi Hospital in Thailand has launched a "PGx" wallet card. The latter summarizes patients' HLA gene variant information to predict the risk of developing SJS/10 (Borchers et al., 2008). Such PGx cards could be readily expandable toward tailor-made therapeutics (Fig. 26.2).

In Latin America, the Iberian American Network of Pharmacogenetics and Pharmacogenomics (RIBEF), created in 2006, aims to promote collaborative PGx research. RIBEF, which consists of 43 research groups and more than 200 researchers, aims to promote scientific studies among its members as well as the clinical implementation of PGx to support the healthcare needs of neglected populations. RIBEF teaching programs and human resources training activities include over 400 events all over Latin America. Moreover, the RIBEF network develops research projects that include Iberoamerican population PGx studies. The Consorcio Europeo e Iberoamericano de Farmacogenética de Poblaciones Consortium (CEIBA) was established among the RIBEF members for this purpose. The MESTIFAR project aimed to determine the variability of polymorphisms in genes involved in the response to drugs in populations of different ethnic origin (Native Americans [Amerindian] and Mestizos [the result of post-Columbian admixture]). In addition to population PGx, RIBEF has projects that relate to clinical PGx in Neurology, Psychiatry, Cardiovascular, and/or Infectious Diseases, resulting in a total of 31 scientific articles being published so far.

In Africa, a disproportionate burden of disease is observed (HIV/AIDS, tuberculosis, and malaria) against a backdrop of an increasing burden of noncommunicable diseases (Niemz et al., 2011). Genomic data have supported the notion that several genetic variants can provide an increased resistance or susceptibility to HIV infections (Sirugo et al., 2008). Notably, a huge variability has been evident regarding the pattern of genetic variations in the CYP genes among African populations. The latter was translated into differential drug responses (Dandara et al., 2014).

In the Middle East, particularly the United Arab Emirates (UAE), PGx research was initially conducted on erythrocyte glucose-6-phosphate dehydrogenase deficiency (G-6-PD) and its association to drug-induced hemolytic anemia (Bayoumi et al., 1996) and later arylamine N-acetyltransferase (NAT2) (Woolhouse et al., 1997). *CYP2D6* allele frequencies were also investigated in the Emirati population, reporting four novel *CYP2D6* variants (Qumsieh et al., 2011), while a warfarin PGx study is under way for the Emirati population (Fortina et al., 2014; Al-Jaibeji et al. unpublished). These studies sparked an interest from Dubai Hospital toward the integration of PGx information for chemotherapeutic agents, while the UAE Health Authority policy of reporting adverse drug reactions in the UAE requires expert pharmacogenomic recommendations within the first 24 h related to each adverse drug reaction reported (Abu Dhabi). PGx awareness is gaining



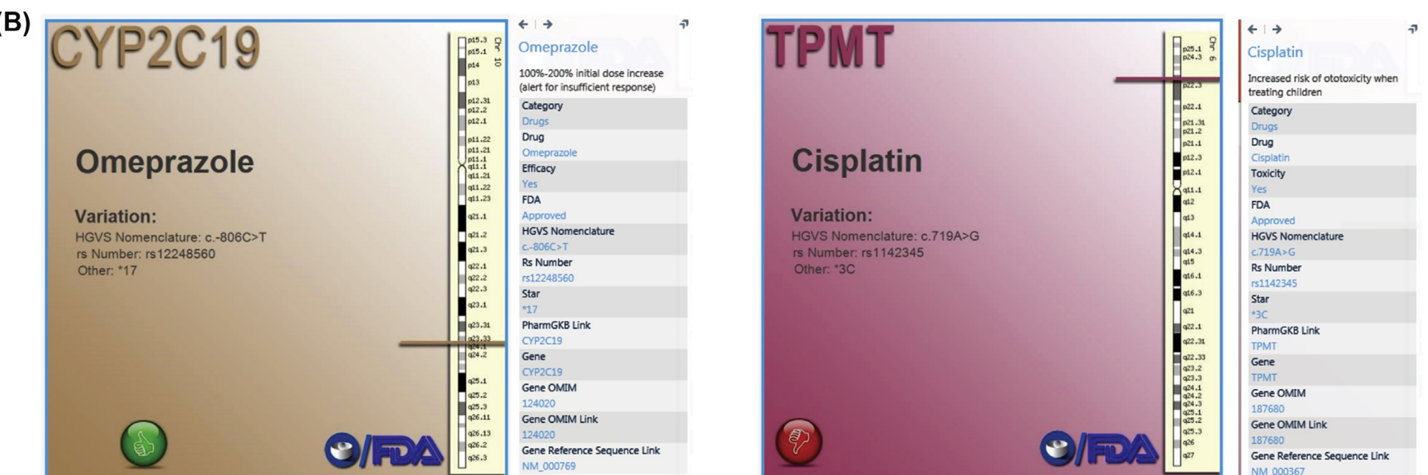


FIGURE 26.1 DruGeVar database. (A) Overview of data entries (colored boxes) on the basis of Microsoft's PivotViewer and Silverlight technologies. The querying interface, by which the user can further exploit the database content, is shown on the left. (B) Display items that correspond to a variant of interest in relation to drug toxicity (*red sign* at the bottom left of the item Dark Gray in print versions) or efficacy (*green sign* at the bottom left of the item Light Gray in print versions).

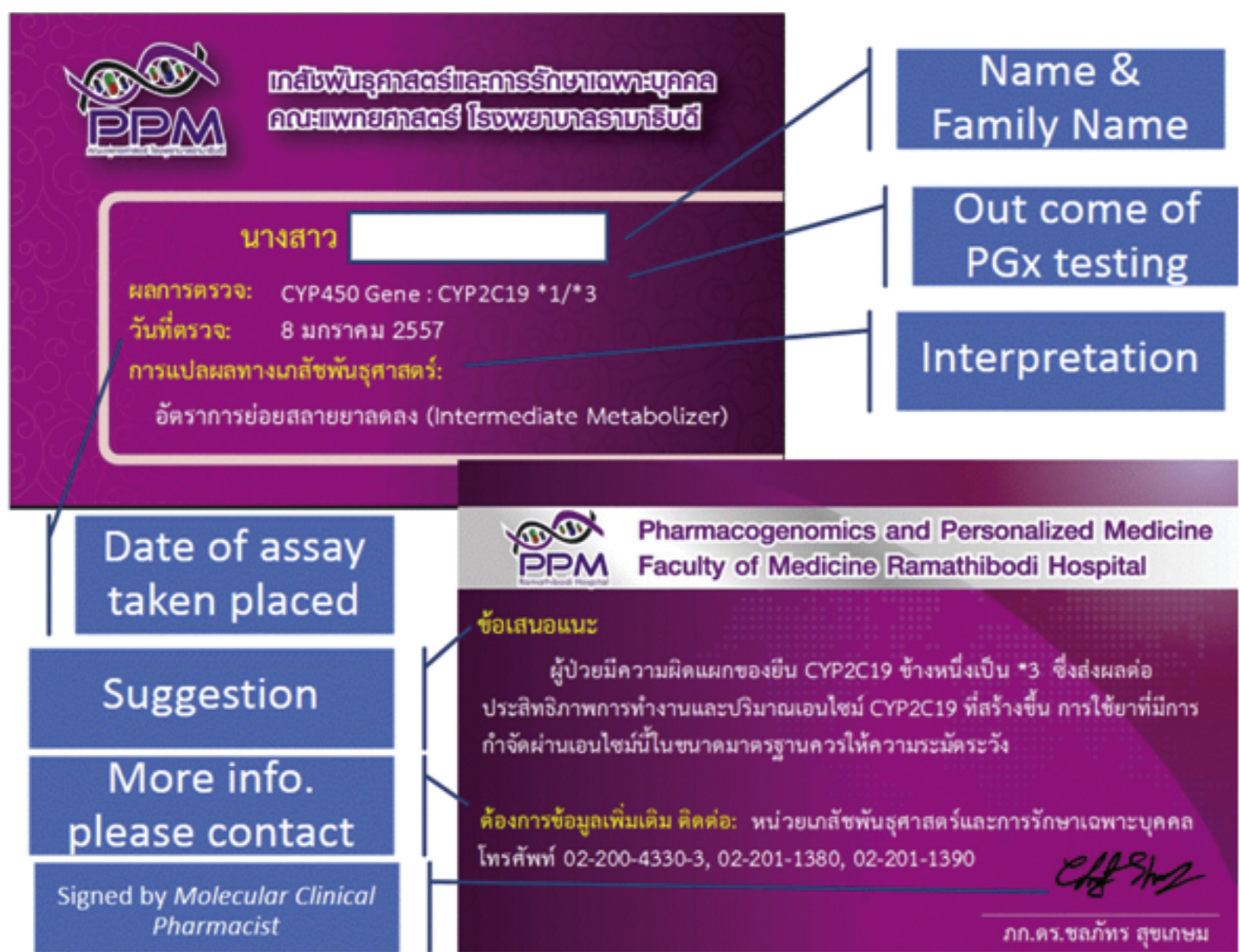


FIGURE 26.2 **"Pharmacogenomics (PGx)" wallet card that assists in clinical decision making.** The card shares information regarding the pharmacogenetics test ordered (*HLA-B* genotyping) and its outcome (*HLA-B*58:01/15:02*). Pharmacogenetics data interpretation informs the clinician about the associated risk of allopurinol, carbamazepine, and oxcarbazepine treatment. Such a pharmacogenomic card has been successfully implemented in clinical practice in Thailand at the Laboratory for PGx, Somdech Phra Debaratana Medical Center, Ramathibodi Hospital.

ground in Middle Eastern countries, such as Saudi Arabia (Abu-Elmagd et al., 2015), Oman (Pathare et al., 2012), Lebanon (Ossaily and Zgheib, 2014), and Qatar (Elewa et al., 2015).

26.1.3 Mapping of Stakeholders in Genomic Medicine

The genomic medicine puzzle is comprised of several key players and stakeholders, and notably, their genomics awareness and views vary significantly. A systematic mapping of those views and different awareness levels would positively impact a better understanding of the policy environment as well as the role of the relevant key stakeholders in the field. Mitropoulou and coworkers undertook such an initiative and assessed the level of support or opposition to PGx and genomic medicine in Greece (Mitropoulou et al., 2014). Similarly, an analysis is currently underway in Middle Eastern countries to determine the stakeholders and their views.

The smooth incorporation of genomic medicine into clinical practice is hindered by insufficient genomics education and a lack of genomics awareness among healthcare professionals and the general public (Reydon et al., 2012). The low genomic literacy of the broader public (and patients) is especially challenging for public health genomics as well as health literacy (Syurina et al., 2011). On top of this, genomics education is not uniformly provided in the various academic institutions worldwide (Pisanu et al., 2014; Mai et al., 2014). Such studies might provide a basis for harmonizing PGx education in southeast European countries with those of northwest European countries to create a smoother and more timely integration of PGx into mainstream medical practice. In Latin America there are

very few postgraduate programs focusing on genomics (Palacios and Collado-Vides, 2007). In Africa, the high cost of genomic services and low private investment are compounded by a relatively low level of medical professionals with an understanding of genomics (Wonkam et al., 2006). An attempt in sub-Saharan Africa to triangulate the views of multiple stakeholders related to prenatal diagnosis of sickle cell disease showed several discrepancies that signal potential value-based conflicts and can usefully inform future policy actions (Wonkam and Hurst, 2014). In fact, African-based scientists are participating in studies focusing on the genomics of monogenic diseases (Mercier et al., 2013; Mtatiro et al., 2014; Wonkam et al., 2014; Wonkam, 2015) and multifactorial conditions (Tekola Ayele et al., 2012). These data are concretely assisting the effective practice of genomic medicine that is well established in South Africa (Beighton et al., 2012), in some Northern African countries (Chaabouni-Bouhamed, 2008; El-Beshlawy et al., 2012), and recently initiated in Central Africa (Wonkam et al., 2011). Regional initiatives, such as the Southern Africa Human Genome Project (Pepper, 2011), have been boosted by international funding agencies and academic institutions through major programs, such as the Malaria Genomic Epidemiology Network (http://www.malariagen.net), the Human Heredity and Health in Africa program (H3Africa Consortium et al., 2014), and the African Genome Variation Project (Gurdasani et al., 2015).

Latin America populations are characterized by high and heterogeneous levels of admixture that arises from their history and corresponds to different patterns of mating between Native American, European, and African individuals (Ruiz-Linares et al., 2014). The National Institute of Genomic Medicine was built in Mexico with public funds (Jimenez-Sanchez et al., 2008), having several publications in the fields of population genomics and medical genomics (Silva-Zolezzi et al., 2009; Moreno-Estrada et al., 2014). Brazil and Colombia have also successfully implemented genomic approaches for the study of several human diseases with a high epidemiological impact in those resource-limited countries (Passos-Bueno et al., 2014; Ortega-Recalde et al., 2014; Pinto et al., 2015; Benitez et al., 2010), while the availability of commercial tests by service providers abroad has been considered to be fundamental toward the implementation of genomic medicine.

26.1.4 Advent of Next-Generation Sequencing in Genomic Medicine

The advent of NGS technology has marked the beginning of a new era in the analysis of human genome sequences (Hodges et al., 2007; Albert et al., 2007; Shendure and Ji, 2008; Gnirke et al., 2009; see also Chapter 9). Before NGS, Sanger sequencing was widely used for screening, even though analyses involving numerous genes or large genomic regions were particularly challenging, e.g., a read length encompassing up to 700 base-pairs (bp) per reaction. The NGS approach allows for the simultaneous analysis of millions of bp in only hours, facilitating large-scale explorations of the human genome (McCarthy et al., 2013). Successful screening attempts focus on researching novel recessive disease-related sequence variants, particularly those caused by homozygous mutations as well as monogenic dominant Mendelian disorder. NGS has not been widely used for some complex pathologies in which several variants might contribute toward the phenotype, because data analysis highlights a remarkable complexity, especially for simultaneous interactive network exploration.

Three main approaches are employed, which depend on the length of the genome region being analyzed: whole-genome sequencing (WGS), whole-exome sequencing (WES), and custom target sequencing microarrays (TSMs) (McCarthy et al., 2013). WGS is mainly used for research, whereas WES and TSM are used for both research and diagnostic purposes. Although most academic and private technological platforms for the previously mentioned are located in high-income countries, resource-limited countries have performed interesting studies by using NGS outsourcing services (Pinto et al., 2015; Benitez et al., 2010; Diggle et al., 2012). Innovative diagnostic approaches have also been proposed for pathologies with overlapping phenotypes caused by several genes (Pinto et al., 2015; Benitez et al., 2010). Knowledge of the genomes of mammalian species has enabled large-scale comparative genomic approaches, resulting in dissecting loci related to evolution mechanisms, which may contribute to human diseases (Prada and Laissue, 2014). Complex pathologies, such as female infertility, have also been explored via NGS (Fonseca et al., 2015).

26.1.5 Is Genomics-Guided Therapy Cost-Effective?

Genome-based drug treatment is expected to reduce national healthcare expenditures. In resource-limited countries in particular, which in many cases have vast fiscal deficits, the economic evaluation of PGx is fundamental (Snyder et al., 2014).

Although the field of economic evaluation in genomic medicine, PGx, and public health genomics is currently in its infancy, several studies indicate that genotype-guided therapy can be cost-effective and of a high cost benefit. Focusing on resource-limited countries, initial economic evaluation studies in the Thai population indicated that HLA-B*1502-guided carbamazepine treatment is cost-effective compared to conventional treatment and can reduce carbamazepine-induced severe adverse drug

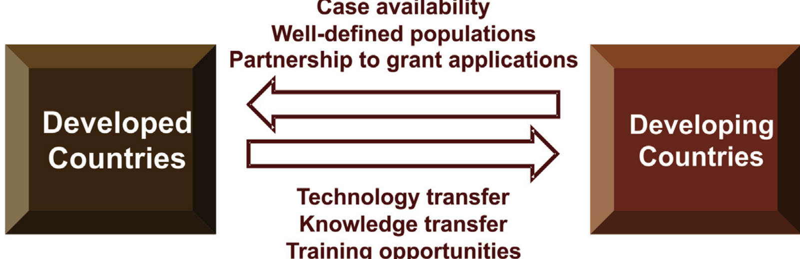


FIGURE 26.3 Encouraging collaboration between developed and developing/resource-limited countries in the field of Genomic Medicine. Developing countries are expected to benefit from training opportunities, knowledge transfer, and/or expanding research networks. Developed countries may also benefit through comparative work as well as multicenter projects on rare diseases or unique clinical features from well-defined populations.

reactions (Rattanavipapong et al., 2013; Grosse, 2008). Similar findings were reported for the Singaporean population, when the cost-effectiveness of HLA-B*1502 genotyping in adult patients with newly diagnosed epilepsy was assessed (Dong et al., 2012).

Another study to evaluate cost-effectiveness of warfarin treatment in Croatian elderly ischemic stroke patients with atrial fibrillation indicated that PGx-guided warfarin treatment represented a cost-effective therapy option for the management of those patients (Mitropoulou et al., 2015).

26.2 CONCLUSIONS AND FUTURE PERSPECTIVES

When resource-limited environments are considered, genomic medicine can only be implemented via a stronger collaboration in genomics research between developed and developing/resource-limited countries, which is likely to create benefits for all parties. Developing countries will benefit from training opportunities, knowledge transfer, and/or expanding transnational networks, and developed countries may benefit through comparative work and multicenter projects on families with rare diseases and/or unique clinical features Fig. 26.3 (Cooper et al., 2014).

Developing countries may suffer from limited resources, but they are also potential-rich in producing data (in the context of genomic medicine-related disciplines, from the perspective of public health genomics). We feel that the success stories presented in this chapter set the paradigm for replication in other countries to acquire more and better insights toward the implementation of genomic medicine and harmonizing the strategies and policies if a fast and smooth adoption of genomic medicine practices occurs in the various national healthcare systems.

ACKNOWLEDGMENTS

This chapter was encouraged by the Genomic Medicine Alliance Public Health Genomics Working Group.

FINANCIAL AND COMPETING INTERESTS

KM is a scientific advisor of the Golden Helix Foundation and GPP is a member of the Scientific Advisory Committee of the Genomic Medicine Alliance. DAF is supported by research grants from Colciencias and VCTI-UAN. The authors declare that they have no competing interests.

REFERENCES

Abu Dhabi, H.A.A., Reporting Adverse Reaction. https://www.haad.ae/haad/Portals/0/ReportingAdverseReactions-updated24-june.pdf .

Abu-Elmagd, M., Assidi, M., Schulten, H.J., Dallol, A., Pushparaj, P., Ahmed, F., Scherer, S.W., Al-Qahtani, M., 2015. Individualized medicine enabled by genomics in Saudi Arabia. BMC Med. Genom. 8 (Suppl. 1), S3.

Albert, T.J., Molla, M.N., Muzny, D.M., Nazareth, L., Wheeler, D., Song, X., Richmond, T.A., Middle, C.M., Rodesch, M.J., Packard, C.J., Weinstock, G.M., Gibbs, R.A., 2007. Direct selection of human genomic loci by microarray hybridization. Nat. Methods 4 (11), 903−905.

Borchers, A.T., Lee, J.L., Naguwa, S.M., Cheema, G.S., Gershwin, M.E., 2008. Stevens-Johnson syndrome and toxic epidermal necrolysis. Autoimmun. Rev. 7 (8), 598−605.

Bayoumi, R.A., Nur-E-Kamal, M.S., Tadayyon, M., Mohamed, K.K., Mahboob, B.H., Qureshi, M.M., Lakhani, M.S., Awaad, M.O., Kaeda, J., Vulliamy, T.J., Luzzatto, L., 1996. Molecular characterization of erythrocyte glucose-6-phosphate dehydrogenase deficiency in Al-Ain District, United Arab Emirates. Hum. Hered. 46 (3), 136−141.

Beighton, P., Fieggen, K., Wonkam, A., Ramesar, R., Greenberg, J., 2012. UCT's contribution to medical genetics in Africa - from the past into the future. S. Afr. Med. J. 102 (6), 446−448.

Benitez, B.A., Forero, D.A., Arboleda, G.H., Granados, L.A., Yunis, J.J., Fernandez, W., Arboleda, H., 2010. Exploration of genetic susceptibility factors for Parkinson's disease in a South American sample. J. Genet. 89 (2), 229−232.

Chung, W.H., Hung, S.I., Hong, H.S., Hsih, M.S., Yang, L.C., Ho, H.C., Wu, J.Y., Chen, Y.T., 2004. Medical genetics: a marker for Stevens-Johnson syndrome. Nature 428 (6982), 486.

Chen, P., Lin, J.J., Lu, C.S., Ong, C.T., Hsieh, P.F., Yang, C.C., Tai, C.T., Wu, S.L., Lu, C.H., Hsu, Y.C., Yu, H.Y., Ro, L.S., Lu, C.T., Chu, C.C., Tsai, J.J., Su, Y.H., Lan, S.H., Sung, S.F., Lin, S.Y., Chuang, H.P., Huang, L.C., Chen, Y.J., Tsai, P.J., Liao, H.T.,

Lin, Y.H., Chen, C.H., Chung, W.H., Hung, S.I., Wu, J.Y., Chang, C.F., Chen, L., Chen, Y.T., Shen, C.Y., Taiwan SJS Consortium, 2011. Carbamazepine-induced toxic effects and HLA-B*1502 screening in Taiwan. N. Engl. J. Med. 364 (12), 1126–1133.

Chaabouni-Bouhamed, H., 2008. Tunisia: communities and community genetics. Community Genet. 11 (6), 313–323.

Cooper, D.N., Brand, A., Dolzan, V., Fortina, P., Innocenti, F., Lee, M.T., Macek Jr., M., Al-Mulla, F., Prainsack, B., Squassina, A., Vayena, E., Vozikis, A., Williams, M.S., Patrinos, G.P., 2014. Bridging genomics research between developed and developing countries: the Genomic Medicine Alliance. Per. Med. 11 (7), 615–623.

Dalabira, E., Viennas, E., Daki, E., Komianou, A., Bartsakoulia, M., Poulas, K., Katsila, T., Tzimas, G., Patrinos, G.P., 2014. DruGeVar: an online resource triangulating drugs with genes and genomic biomarkers for clinical pharmacogenomics. Public Health Genom. 17 (5–6), 265–271.

Dandara, C., Swart, M., Mpeta, B., Wonkam, A., Masimirembwa, C., 2014. Cytochrome P450 pharmacogenetics in African populations: implications for public health. Expert Opin. Drug Metab. Toxicol. 10 (6), 769–785.

Diggle, C.P., Parry, D.A., Logan, C.V., Laissue, P., Rivera, C., Restrepo, C.M., Fonseca, D.J., Morgan, J.E., Allanore, Y., Fontenay, M., Wipff, J., Varret, M., Gibault, L., Dalantaeva, N., Korbonits, M., Zhou, B., Yuan, G., Harifi, G., Cefle, K., Palanduz, S., Akoglu, H., Zwijnenburg, P.J., Lichtenbelt, K.D., Aubry-Rozier, B., Superti-Furga, A., Dallapiccola, B., Accadia, M., Brancati, F., Sheridan, E.G., Taylor, G.R., Carr, I.M., Johnson, C.A., Markham, A.F., Bonthron, D.T., 2012. Prostaglandin transporter mutations cause pachydermoperiostosis with myelofibrosis. Hum. Mutat. 33 (8), 1175–1181.

Dong, D., Sung, C., Finkelstein, E.A., 2012. Cost-effectiveness of HLA-B*1502 genotyping in adult patients with newly diagnosed epilepsy in Singapore. Neurology 79 (12), 1259–1267.

Elewa, H., Alkhiyami, D., Alsahan, D., Abdel-Aziz, A., August 2015. A survey on the awareness and attitude of pharmacists and doctors towards the application of pharmacogenomics and its challenges in Qatar. J. Eval. Clin. Pract. 21 (4), 703–709.

El-Beshlawy, A., El-Shekha, A., Momtaz, M., Said, F., Hamdy, M., Osman, O., et al., 2012. Prenatal diagnosis for thalassaemia in Egypt: what changed parents' attitude? Prenat. Diagn. 32 (8), 777–782.

Fortina, P., Al Khaja, N., Al Ali, M.T., Hamzeh, A.R., Nair, P., Innocenti, F., Patrinos, G.P., Kricka, L.J., 2014. Genomics into healthcare: the 5th Pan Arab Human Genetics Conference and 2013 Golden Helix Symposium. Hum. Mutat. 35 (5), 637–640.

Fonseca, D.J., Patiño, L.C., Suárez, Y.C., de Jesús Rodríguez, A., Mateus, H.E., Jiménez, K.M., Ortega-Recalde, O., Díaz-Yamal, I., Laissue, P., July 2015. Next generation sequencing in women affected by nonsyndromic premature ovarian failure displays new potential causative genes and mutations. Fertil. Steril. 104 (1), 154–162 e2.

Giardine, B., Borg, J., Higgs, D.R., Peterson, K.R., Philipsen, S., Maglott, D., Singleton, B.K., Anstee, D.J., Basak, A.N., Clark, B., Costa, F.C., Faustino, P., Fedosyuk, H., Felice, A.E., Francina, A., Galanello, R., Gallivan, M.V., Georgitsi, M., Gibbons, R.J., Giordano, P.C., Harteveld, C.L., Hoyer, J.D., Jarvis, M., Joly, P., Kanavakis, E., Kollia, P., Menzel, S., Miller, W., Moradkhani, K., Old, J., Papachatzopoulou, A., Papadakis, M.N., Papadopoulos, P., Pavlovic, S., Perseu, L., Radmilovic, M., Riemer, C., Satta, S., Schrijver, I., Stojiljkovic, M., Thein, S.L., Traeger-Synodinos, J., Tully, R., Wada, T., Waye, J.S., Wiemann, C., Zukic, B., Chui, D.H., Wajcman, H., Hardison, R.C., Patrinos, G.P., 2011. Systematic documentation and analysis of human genetic variation in hemoglobinopathies using the microattribution approach. Nat. Genet. 43 (4), 295–301.

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M.O., Choudhury, A., Ritchie, G.R., Xue, Y., Asimit, J., Nsubuga, R.N., Young, E.H.1, Pomilla, C.1, Kivinen, K., Rockett, K., Kamali, A., Doumatey, A.P., Asiki, G., Seeley, J., Sisay-Joof, F., Jallow, M., Tollman, S., Mekonnen, E., Ekong, R., Oljira, T., Bradman, N., Bojang, K., Ramsay, M., Adeyemo, A., Bekele, E., Motala, A., Norris, S.A., Pirie, F., Kaleebu, P., Kwiatkowski, D., Tyler-Smith, C., Rotimi, C., Zeggini, E., Sandhu, M.S., 2015. The African genome variation project shapes medical genetics in Africa. Nature 517 (7534), 327–332.

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E.M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D.B., Lander, E.S., Nusbaum, C., 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat. Biotechnol. 27 (2), 182–189.

Grosse, S.D., 2008. Assessing cost-effectiveness in healthcare: history of the $50,000 per QALY threshold. Expert Rev. Pharmacoecon. Outcomes Res. 8 (2), 165–178.

Hung, S.I., Chung, W.H., Liou, L.B., Chu, C.C., Lin, M., Huang, H.P., Lin, Y.L., Lan, J.L., Yang, L.C., Hong, H.S., Chen, M.J., Lai, P.C., Wu, M.S., Chu, C.Y., Wang, K.H., Chen, C.H., Fann, C.S., Wu, J.Y., Chen, Y.T., 2005. HLA-B*5801 allele as a genetic marker for severe cutaneous adverse reactions caused by allopurinol. Proc. Natl. Acad. Sci. U. S. A. 102 (11), 4134–4139.

H3Africa Consortium, Rotimi, C., Abayomi, A., Abimiku, A., Adabayeri, V.M., Adebamowo, C., Adebiyi, E., Ademola, A.D., Adeyemo, A., Adu, D., Affolabi, D., Agongo, G., Ajayi, S., Akarolo-Anthony, S., Akinyemi, R., Akpalu, A., Alberts, M., Alonso Betancourt, O., Alzohairy, A.M., Ameni, G., Amodu, O., Anabwani, G., Andersen, K., Arogundade, F., Arulogun, O., Asogun, D., Bakare, R., Balde, N., Baniecki, M.L., Beiswanger, C., Benkahla, A., Bethke, L., Boehnke, M., Boima, V., Brandful, J., Brooks, A.I., Brosius, F.C., Brown, C., Bucheton, B., Burke, D.T., Burnett, B.G., Carrington-Lawrence, S., Carstens, N., Chisi, J., Christoffels, A., Cooper, R., Cordell, H., Crowther, N., Croxton, T., de Vries, J., Derr, L., Donkor, P., Doumbia, S., Duncanson, A., Ekem, I., El Sayed, A., Engel, M.E., Enyaru, J.C., Everett, D., Fadlelmola, F.M., Fakunle, E., Fischbeck, K.H., Fischer, A., Folarin, O., Gamieldien, J., Garry, R.F., Gaseitsiwe, S., Gbadegesin, R., Ghansah, A., Giovanni, M., Goesbeck, P., Gomez-Olive, F.X., Grant, D.S., Grewal, R., Guyer, M., Hanchard, N.A., Happi, C.T., Hazelhurst, S., Hennig, B.J., Hertz- C, F., Hide, W., Hilderbrandt, F., Hugo-Hamman, C., Ibrahim, M.E., James, R., Jaufeerally-Fakim, Y., Jenkins, C., Jentsch, U., Jiang, P.P., Joloba, M., Jongeneel, V., Joubert, F., Kader, M., Kahn, K., Kaleebu, P., Kapiga, S.H., Kassim, S.K., Kasvosve, I., Kayondo, J., Keavney, B., Kekitiinwa, A., Khan, S.H., Kimmel, P., King, M.C., Kleta, R., Koffi, M., Kopp, J., Kretzler, M., Kumuthini, J., Kyobe, S.,

Kyobutungi, C., Lackland, D.T., Lacourciere, K.A., Landouré, G., Lawlor, R., Lehner, T., Lesosky, M., Levitt, N., Littler, K., Lombard, Z., Loring, J.F., Lyantagaye, S., Macleod, A., Madden, E.B., Mahomva, C.R., Makani, J., Mamven, M., Marape, M., Mardon, G., Marshall, P., Martin, D.P., Masiga, D., Mason, R., Mate-Kole, M., Matovu, E., Mayige, M., Mayosi, B.M., Mbanya, J.C., McCurdy, S.A., McCarthy, M.I., McIlleron, H., Mc'Ligeyo, S.O., Merle, C., Mocumbi, A.O., Mondo, C., Moran, J.V., Motala, A., Moxey-Mims, M., Mpoloka, W.S., Msefula, C.L., Mthiyane, T., Mulder, N., Mulugeta, G., Mumba, D., Musuku, J., Nagdee, M., Nash, O., Ndiaye, D., Nguyen, A.Q., Nicol, M., Nkomazana, O., Norris, S., Nsangi, B., Nyarko, A., Nyirenda, M., Obe, E., Obiakor, R., Oduro, A., Ofori-Acquah, S.F., Ogah, O., Ogendo, S., Ohene-Frempong, K., Ojo, A., Olanrewaju, T., Oli, J., Osafo, C., Ouwe Missi Oukem-Boyer, O., Ovbiagele, B., Owen, A., Owolabi, M.O., Owolabi, L., Owusu-Dabo, E., Pare, G., Parekh, R., Patterton, H.G., Penno, M.B., Peterson, J., Pieper, R., Plange-Rhule, J., Pollak, M., Puzak, J., Ramesar, R.S., Ramsay, M., Rasooly, R., Reddy, S., Sabeti, P.C., Sagoe, K., Salako, T., Samassékou, O., Sandhu, M.S., Sankoh, O., Sarfo, F.S., Sarr, M., Shaboodien, G., Sidibe, I., Simo, G., Simuunza, M., Smeeth, L., Sobngwi, E., Soodyall, H., Sorgho, H., Sow Bah, O., Srinivasan, S., Stein, D.J., Susser, E.S., Swanepoel, C., Tangwa, G., Tareila, A., Tastan Bishop, O., Tayo, B., Tiffin, N., Tinto, H., Tobin, E., Tollman, S.M., Traoré, M., Treadwell, M.J., Troyer, J., Tsimako-Johnstone, M., Tukei, V., Ulasi, I., Ulenga, N., van Rooyen, B., Wachinou, A.P., Waddy, S.P., Wade, A., Wayengera, M., Whitworth, J., Wideroff, L., Winkler, C.A., Winnicki, S., Wonkam, A., Yewondwos, M., sen, T., Yozwiak, N., Zar, H., 2014. Research capacity. Enabling the genomic revolution in Africa. Science 344 (6190), 1346–1348.

Hodges, E., Xuan, Z., Balija, V., Kramer, M., Molla, M.N., Smith, S.W., Middle, C.M., Rodesch, M.J., Albert, T.J., Hannon, G.J., McCombie, W.R., 2007. Genome-wide in situ exon capture for selective resequencing. Nat. Genet. 39 (12), 1522–1527.

Jimenez-Sanchez, G., Silva-Zolezzi, I., Hidalgo, A., March, S., 2008. Genomic medicine in Mexico: initial steps and the road ahead. Genome Res. 18 (8), 1191–1198.

McCarthy, J.J., McLeod, H.L., Ginsburg, G.S., 2013. Genomic medicine: a decade of successes, challenges, and opportunities. Sci. Transl. Med. 5 (189), 189sr4.

Mardis, E.R., 2011. A decade's perspective on DNA sequencing technology. Nature 470 (7333), 198–203.

Mitropoulos, K., Johnson, L., Vozikis, A., Patrinos, G.P., 2011. Relevance of pharmacogenomics for developing countries in Europe. Drug Metabol. Drug Interact. 26 (4), 143–146.

Mitropoulou, C., Mai, Y., van Schaik, R.H., Vozikis, A., Patrinos, G.P., 2014. Stakeholder analysis in pharmacogenomics and genomic medicine in Greece. Public Health Genom. 17 (5–6), 280–286.

Mai, Y., Mitropoulou, C., Papadopoulou, X.E., Vozikis, A., Cooper, D.N., van Schaik, R.H., Patrinos, G.P., 2014. Critical appraisal of the views of healthcare professionals with respect to pharmacogenomics and personalized medicine in Greece. Pers. Med. 11 (1), 15–26.

Mercier, S., Küry, S., Shaboodien, G., Houniet, D.T., Khumalo, N.P., Bou-Hanna, C., Bodak, N., Cormier-Daire, V., David, A., Faivre, L., Figarella-Branger, D., Gherardi, R.K., Glen, E., Hamel, A., Laboisse, C., Le Caignec, C., Lindenbaum, P., Magot, A., Munnich, A., Mussini, J.M., Pillay, K., Rahman, T., Redon, R., Salort-Campana, E., Santibanez-Koref, M., Thauvin, C., Barbarot, S., Keavney, B., Bézieau, S., Mayosi, B.M., 2013. Mutations in *FAM111B* cause hereditary fibrosing poikiloderma with tendon contracture, myopathy, and pulmonary fibrosis. Am. J. Hum. Genet. 93 (6), 1100–1107.

Mtatiro, S.N., Singh, T., Rooks, H., Mgaya, J., Mariki, H., Soka, D., Mmbando, B., Msaki, E., Kolder, I., Thein, S.L., Menzel, S., Cox, S.E., Makani, J., Barrett, J.C., 2014. Genome wide association study of fetal hemoglobin in sickle cell anemia in Tanzania. PLoS One 9 (11), e111464.

Moreno-Estrada, A., Gignoux, C.R., Fernández-López, J.C., Zakharia, F., Sikora, M., Contreras, A.V., Acuña-Alonzo, V., Sandoval, K., Eng, C., Romero-Hidalgo, S., Ortiz-Tello, P.4, Robles, V., Kenny, E.E., Nuño-Arana, I., Barquera-Lozano, R., Macín-Pérez, G., Granados-Arriola, J., Huntsman, S., Galanter, J.M., Via, M., Ford, J.G., Chapela, R.2, Rodriguez-Cintron, W., Rodríguez-Santana, J.R., Romieu, I., Sienra-Monge, J.J., del Rio Navarro, B., London, S.J., Ruiz-Linares, A., Garcia-Herrera, R., Estrada, K., Hidalgo-Miranda, A., Jimenez-Sanchez, G., Carnevale, A., Soberón, X., Canizales-Quinteros, S., Rangel-Villalobos, H., Silva-Zolezzi, I., Burchard, E.G., Bustamante, C.D., 2014. Human genetics. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. Science 344 (6189), 1280–1285.

Mitropoulou, C., Fragoulakis, V., Bozina, N., Vozikis, A., Supe, S., Bozina, T., Poljakovic, Z., van Schaik, R.H., Patrinos, G.P., 2015. Economic evaluation of pharmacogenomic-guided warfarin treatment for elderly Croatian atrial fibrillation patients with ischemic stroke. Pharmacogenomics 16 (2), 137–148.

Niemz, A., Ferguson, T.M., Boyle, D.S., 2011. Point-of-care nucleic acid testing for infectious diseases. Trends Biotechnol. 29 (5), 240–250.

Ossaily, S., Zgheib, N.K., 2014. The pharmacogenetics of drug metabolizing enzymes in the Lebanese population. Drug Metabol. Drug Interact. 29 (2), 81–90.

Ortega-Recalde, O., Vergara, J.I., Fonseca, D.J., Ríos, X., Mosquera, H., Bermúdez, O.M., Medina, C.L., Vargas, C.I., Pallares, A.E., Restrepo, C.M., Laissue, P., 2014. Whole-exome sequencing enables rapid determination of xeroderma pigmentosum molecular etiology. PLoS One 8 (6), e64692.

Pathare, A.V., Al Zadjali, S., Misquith, R., Alkindi, S.S., Panjwani, V., Lapoumeroulie, C., Pravin, S., Paldi, A., Krishnamoorthy, R., February 2012. Warfarin pharmacogenetics: polymorphisms of the CYP2C9, CYP4F2, and VKORC1 loci in a genetically admixed Omani population. Hum. Biol. 84 (1), 67–77.

Pisanu, C., Tsermpini, E.E., Mavroidi, E., Katsila, T., Patrinos, G.P., Squassina, A., 2014. Assessment of the pharmacogenomics educational environment in Southeast Europe. Public Health Genom. 17 (5–6), 272–279.

Palacios, R., Collado-Vides, J., 2007. Development of genomic sciences in Mexico: a good start and a long way to go. PLoS Comput. Biol. 3 (9), 1670–1673.

Pepper, M.S., 2011. Launch of the Southern African Human Genome Programme. South Afr. Med. J. 101 (5), 287–288.

Passos-Bueno, M.R., Bertola, D., Horovitz, D.D., de Faria Ferraz, V.E., Brito, L.A., 2014. Genetics and genomics in Brazil: a promising future. Mol. Genet. Genomic Med. 2 (4), 280–291.

Pinto, W.B., Oliveira, A.S., Souza, P.V., 2015. Whole exome sequencing identifies three recessive *FIG*$_4$-mutations in an

apparently dominant pedigree with Charcot-Marie-Tooth disease. Neuromuscul. Disord. 25 (4), 359−360.

Prada, C.F., Laissue, P., 2014. A high resolution map of mammalian X chromosome fragile regions assessed by large-scale comparative genomics. Mamm. Genome 25 (11−12), 618−635.

Qumsieh, R.Y., Ali, B.R., Abdulrazzaq, Y.M., Osman, O., Akawi, N.A., Bastaki, S.M., 2011. Identification of new alleles and the determination of alleles and genotypes frequencies at the *CYP2D6* gene in Emiratis. PLoS One 6 (12), e28943.

Reydon, T.A., Kampourakis, K., Patrinos, G.P., 2012. Genetics, genomics and society: the responsibilities of scientists for science communication and education. Pers. Med. 9 (6), 633−643.

Ruiz-Linares, A., Adhikari, K., Acuña-Alonzo, V., Quinto-Sanchez, M., Jaramillo, C., Arias, W., Fuentes, M., Pizarro, M., Everardo, P., de Avila, F., Gómez-Valdés, J., León-Mimila, P., Hunemeier, T., Ramallo, V., Silva de Cerqueira, C.C., Burley, M.W., Konca, E., de Oliveira, M.Z., Veronez, M.R., Rubio-Codina, M., Attanasio, O., Gibbon, S., Ray, N., Gallo, C., Poletti, G., Rosique, J., Schuler-Faccini, L., Salzano, F.M., Bortolini, M.C., Canizales-Quinteros, S., Rothhammer, F., Bedoya, G., Balding, D., Gonzalez-José, R., 2014. Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. PLoS Genet. 10 (9), e1004572.

Rattanavipapong, W., Koopitakkajorn, T., Praditsitthikorn, N., Mahasirimongkol, S., Teerawattananon, Y., 2013. Economic evaluation of HLA-B*15:02 screening for carbamazepine-induced severe adverse drug reactions in Thailand. Epilepsia 54 (9), 1628−1638.

Sirugo, G., Hennig, B.J., Adeyemo, A.A., Matimba, A., Newport, M.J., Ibrahim, M.E., Ryckman, K.K., Tacconelli, A., Mariani-Costantini, R., Novelli, G., Soodyall, H., Rotimi, C.N., Ramesar, R.S., Tishkoff, S.A., Williams, S.M., 2008. Genetic studies of African populations: an overview on disease susceptibility and response to vaccines and therapeutics. Hum. Genet. 123 (6), 557−598.

Syurina, E., Brankovic, I., Probst-Hensch, N., Brand, A., 2011. Genome-based health literacy: a new challenge for public health genomics. Public Health Genom. 14 (4−5), 201−210.

Silva-Zolezzi, I., Hidalgo-Miranda, A., Estrada-Gil, J., Fernandez-Lopez, J.C., Uribe-Figueroa, L., Contreras, A., Balam-Ortiz, E., del Bosque-Plata, L., Velazquez-Fernandez, D., Lara, C., Goya, R., Hernandez-Lemus, E., Davila, C., Barrientos, E., March, S., Jimenez-Sanchez, G., 2009. Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico. Proc. Natl. Acad. Sci. U. S. A. 106 (21), 8611−8616.

Shendure, J., Ji, H., 2008. Next-generation DNA sequencing. Nat. Biotechnol. 26 (10), 1135−1145.

Snyder, S.R., Mitropoulou, C., Patrinos, G.P., Williams, M.S., 2014. Economic evaluation of pharmacogenomics: a value-based approach to pragmatic decision making in the face of complexity. Public Health Genom. 17 (5−6), 256−264.

Tekola Ayele, F., Adeyemo, A., Finan, C., Hailu, E., Sinnott, P., Burlinson, N.D., Aseffa, A., Rotimi, C.N., Newport, M.J., Davey, G., 2012. HLA class II locus and susceptibility to podoconiosis. N. Engl. J. Med. 366 (13), 1200−1208.

Woolhouse, N.M., Qureshi, M.M., Bastaki, S.M., Patel, M., Abdulrazzaq, Y., Bayoumi, R.A., 1997. Polymorphic N-acetyltransferase (NAT2) genotyping of Emiratis. Pharmacogenetics 7 (1), 73−82.

Wonkam, A., Njamnshi, A.K., Angwafo 3rd, F.F., 2006. Knowledge and attitudes concerning medical genetics amongst physicians and medical students in Cameroon (sub-Saharan Africa). Genet. Med. 8 (6), 331−338.

Wonkam, A., Hurst, S., 2014. A call for policy action in sub-Saharan Africa to rethink diagnostics for pregnancy affected by sickle cell disease: differential views of medical doctors, parents and adult patients predict value conflicts in Cameroon. Omics J. Integr. Biol. 18 (7), 472−480.

Wonkam, A., Ngo Bitoungui, V.J., Vorster, A.A., Ramesar, R., Cooper, R.S., Tayo, B., Lettre, G., Ngogang, J., 2014. Association of variants at *BCL11A* and *HBS1L-MYB* with hemoglobin F and hospitalization rates among sickle cell patients in Cameroon. PLoS One 9 (3), e92506.

Wonkam, A., 2015. Letter to the editor regarding "*GJB2, GJB6* or *GJA1* genes should not be investigated in routine in non syndromic deafness in people of sub-Saharan African descent". Int. J. Pediatr. otorhinolaryngology 79 (4), 632−633.

Wonkam, A., Tekendo, C.N., Sama, D.J., Zambo, H., Dahoun, S., Béna, F., Morris, M.A., 2011. Initiation of a medical genetics service in sub-Saharan Africa: experience of prenatal diagnosis in Cameroon. Eur. J. Med. Genet. 54 (4), e399−404.

Chapter 27

Public Understanding of Genetic Testing and Obstacles to Genetics Literacy

K. Kampourakis
University of Geneva, Geneva, Switzerland

27.1 GENETICS LITERACY AND THE PUBLIC UNDERSTANDING OF GENETIC TESTING

27.1.1 Defining Genetics Literacy

Education policy documents in all industrialized societies consider science literacy to be a main goal for education. For instance, the science standards in several European Union member states (Eurydice Network, 2011) and the United States (National Research Council, 2012) suggest that future citizens literate about science, need to acquire the necessary scientific knowledge, as well as to be able to understand the challenges of science in society and the interaction between them. Literacy about genetics is a demand of our times because contemporary genetics research raises a growing number of socio-scientific issues regarding our understanding of health, behavior, and identity. Therefore, a main goal for science education should be to educate future citizens who will be literate about genetics.[1] But what exactly does literacy about genetics consists of? Roughly put, people need to develop two distinct competencies for this purpose: (1) They need to understand how genetics research is conducted and what kind of knowledge it produces, and to be able to use this knowledge in explanations; and (2) they need to develop argumentation skills to be able to argue for and make decisions about socio-scientific issues relevant to genetics (based on Roberts, 2007, pp. 729–730). Thus, literacy about genetics includes knowledge about DNA, genes, chromosomes, and patterns of inheritance, as well as about the process and methods of science, etc.; it also includes knowledge about argumentation and decision making relevant to ethical questions related to all kinds of genetic diagnosis, as well as to information about individuals and privacy issues.

These two competencies may seem to be distinct in principle, but at least in the case of genetics they are related. Research in science education seems to suggest that the ability to engage in reasoning about socio-scientific issues relevant to genetics and gene technologies depends on understanding the relevant science content. For example, in a study with undergraduate students it was found that those with a more advanced understanding of genetics exhibited higher-quality reasoning about gene therapy and cloning scenarios and were more likely to incorporate content knowledge in their reasoning, compared with participants with a lesser understanding of genetics (Sadler and Zeidler, 2005). In another study with junior secondary students, it was found that the ability to engage in reasoning about applications of gene technology (gene transfer, genetic screening, and patenting) was strongly influenced by the ability to recognize key issues, which in turn required some understanding of the relevant science (Lewis and Leach, 2006). Finally, in a study that investigated how high school students, college nonscience majors, and college science majors with advanced genetics knowledge applied genetics content knowledge in justifying claims relative to gene therapy and cloning, science majors did better than the other groups in terms of the quality of their justifications, which were in turn based on specific science content knowledge (Sadler and Fowler, 2006).

These results may not suggest that genetics content knowledge is a sufficient condition for reasoned discussion on socio-scientific issues relevant to genetics, yet they suggest that it is a necessary one. In other words, to make

1. In this chapter, the term "genetics" is used loosely to refer to all kinds of research on biological inheritance, including genomics.

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00027-4
Copyright © 2017 Elsevier Ltd. All rights reserved.

informed decisions about genetic testing and the use of data collected through that, it is important that people become aware of what kind of data are collected, how they can be used, what problems may occur along the way, and what conclusions can be drawn from them. As discussed in more detail subsequently, people often have unreasonable expectations about genetic testing and its conclusions. For example, in the case of forensic DNA evidence, a myth about the infallibility of this method has become prevalent despite various cases of serious mistakes and wrong conclusions (Thompson, 2013). Therefore, it is necessary to examine the quality of understanding of genetics among the public to evaluate whether people have the necessary (but not sufficient) conditions to develop literacy about genetics. Although members of the public should be able to rely on the advice of health care professionals, it is also necessary for them to be able to understand what conclusions can and cannot be drawn from genetics tests.

27.1.2 Public Understanding of Genetic Testing

Reviews of studies on the public understanding of and attitudes toward genetics have revealed several problems. Overall, the general public in so-called developed countries seems to have a relatively high awareness of genetics. In particular, several studies have concluded that most people are familiar with core concepts of genetics, such as genes and heredity, and that they understand that human diseases have both genetic and other causes. However, it seems that people do not have a clear and stable understanding of gene–environment interactions, that they have little knowledge regarding the details of how genes are implicated in the development of traits or of genetic testing procedures, and that they rely on their experiences and models of social relationships to estimate outcomes. Furthermore, whereas it seems that attitudes toward genetic testing for predisposition to disease are highly favorable, these arise in part from the belief that testing allows for the control of health. People have the expectation that genetic tests may provide definite answers, e.g., that one is or is not at risk for a particular condition, whereas they also conflate other kinds of testing with genetic testing. However, interest in genetic testing decreases when the probabilities of having the disease associated with the allele are less than 100% and treatment for the disease is not available. Finally, only a few studies have addressed in detail the public understanding of and attitudes toward the risks associated with genetic testing, with mixed results in studies on more personal risks of genetic testing relevant to autonomy or privacy, and in studies addressing testing in clinical rather than research contexts (Condit, 2010a,b).

Another interesting finding is related to understanding the risks of developing cancer, which is certainly a genetic but not necessarily an inherited disease. In May 2013, actress Angelina Jolie revealed that she had undergone a double mastectomy for preventive purposes because she had been found to carry a rare mutation of the *BRCA1* gene and because she also had a family history of cancer. A survey was conducted in the United States with a representative sample of 2572 adults, who identified their sources of information for this story and also described their understanding of and reaction to it. Participants were asked hypothetical questions regarding preventive surgery in the case of breast cancer, while information about family risk for breast and ovarian cancer was collected. Results indicated that whereas about 74% of participants were aware of the Angelina Jolie case and about 48% could recall Jolie's estimated risk of breast cancer before surgery, fewer than 8% had the necessary information to accurately estimate her risk of developing breast cancer compared with a woman who did not carry the particular *BRCA1* mutation. It was concluded that participants in this study considered a negative family history to be protective against cancer, and that the Angelina Jolie story appeared to confuse them about the relationship between a positive family history and increased cancer risk (Borzekowski et al., 2013).

Not surprisingly, the public also does not seem to be appropriately informed about the applications of direct-to-consumer genetic (DTCG) tests and the conclusions that can be drawn from them. A review of 17 studies on users' perspectives on DTCG tests revealed a low level of awareness as well as concerns about the reliability of these tests and privacy issues. It should be noted, though, that only 2 of the 17 studies reviewed involved actual users of such tests. Participants in the reviewed studies raised two main concerns: privacy issues and the nature of the results and their impact (Goldsmith et al., 2012). Another study that compared how 145 individuals from the general public and 171 genetic counselors interpreted results from DTCG tests suggests that there are important differences between them and that the general public may misinterpret the results of these tests. This study also suggests that at least some members of the public will need guidance to understand the conclusions that can be drawn from their DTCG test results (Leighton et al., 2012). Therefore, as these genetic tests are widely available, the public should be appropriately informed about their applications, the conclusions that can be drawn from them, and the relevant ethical issues (Kampourakis et al., 2014). It should not be assumed that the public could interpret test results accurately.

This is important to keep in mind because the DTCG companies themselves do not provide all of the information necessary to potential users. An analysis of 29 health-related direct-to-consumer websites examined whether their informational content (e.g., benefits, limitations), literacy demands (e.g., reading level), and usability (e.g., ease of navigation) met existing recommendations. The main conclusion was that most sites provided information about the health conditions and the markers for which they tested, the benefits of testing, the testing

process, and their privacy policy. Most sites also listed at least one benefit for consumers undergoing testing, the most common of which was that the test results can inform a health decision. However, only 11 websites provided any scientific evidence in support of their tests; furthermore, only 6 of those cited the relevant scientific literature. Regarding limitations, about half of the websites presented at least one limitation related to testing, the most common one being that other factors besides genes are important for estimating disease risk. Most interestingly, about half of the websites mentioned that the science underlying the tests is new and/or changing, but this was not considered to be a limitation of the tests. Overall, it seems that many users would struggle to find the important information, understand it, and interpret the results of the respective tests appropriately (Lachance et al., 2010).

27.1.3 Health Care Professionals' Understanding of Genetic Testing

It should not be assumed that physicians and other health care professionals could effectively advise the public about these tests. For instance, a national US survey of 10,303 physicians, a sample considered representative of the overall US physician population, indicated that only 29.0% of them had received any education in pharmacogenomics and only 10.3% felt adequately informed about pharmacogenomic testing. In addition, less than 40% of them had recently ordered a test or anticipated ordering a test soon. What these physicians seemed to lack was knowledge about what tests are available, when to use them, where to obtain them, how to interpret the results, and how to apply them in individual cases (Stanek et al., 2012). Another study involved 597 primary care physicians in the United States. The assumption behind this study was that pharmacogenomics tests are primarily used by clinical specialists. However, as the number of drugs for which testing becomes available increases, more and more primary care physicians will have to use such tests. It was found that only 13% of primary care US physicians indicated that they felt comfortable ordering such tests, and 16% reported that they had received some formal education about pharmacogenomics. In addition, 20% had ordered a relevant test at least once a year in the past (Haga et al., 2012). Finally, and perhaps most interestingly, a study with 516 clinical geneticists and genetic counselors indicated that even they might not be prepared for pharmacogenomics counseling. Twelve percent of genetic counselors and 41% of clinical geneticists indicated that they had ordered or coordinated patient care for pharmacogenomics testing. However, and although almost all respondents had some education in pharmacogenomics, only 28% of counselors and 58% of clinical geneticists indicated that they felt well-informed (Haga et al., 2012).

These findings have serious implications for the education of physicians and geneticists who have an important role in pharmacogenomics testing. To overcome the lack of the necessary genetics education, health care professionals should be given opportunities to acquire a better understanding of the relevant issues. Recommendations and ideas for this purpose are already available: (1) increasing the amount of content related to genetics and common diseases as opposed to rare Mendelian diseases in the preservice curriculum; (2) ensure that instruction is case-based and reflects practical examples that demonstrate that genetics matters on a daily basis and can improve patient outcomes; and (3) develop continuing education programs (Guttmacher et al., 2007). That would be a first step. However, in the long run, undergraduate medical training could be revised to include information about recent advances in genetics, whereas professional societies and health care companies could have a positive role in providing physicians with the information and guidelines they need. Although it is important to train health care professionals appropriately, we should not rely on that, but try to support the public to develop the necessary genetics literacy. However, it seems that there are a few crucial obstacles we need to overcome to achieve this.

27.2 OBSTACLES TO GENETICS LITERACY AND HOW THESE MIGHT BE OVERCOME

27.2.1 Genetic Determinism: There Are No "Genes for Traits"

A major issue when it comes to understanding genetics is how people conceive of the roles of DNA and genes. There is a wrong view, genetic determinism, that seems to be widespread at least in formal education, and that constantly refers to the existence of "genes for traits" (this may not be the case in mass media portrayals of genetics) (Condit et al., 1998; Nelkin and Lindee, 2004). According to genetic determinism, genes invariably determine characters and disease, so that the outcomes are just a little or not at all affected by changes in the environment or by the different environments in which an individual lives (Beckwith, 2002; Kitcher, 2003). There are several plausible reasons why genetic determinism is so prevalent. Some of these are school courses in which teachers consistently teach that genes determine some characteristics, mass media headlines about discoveries of genes for traits, and personal observations of the development of characteristics that do not seem to be affected by the environment (Moore, 2008).

Genetic determinism is a serious obstacle to a proper understanding of genetics and its implications for society (Jimenez, 2014). A comparative study of biology textbooks from six countries has shown that they present genetics

concepts in a way that does not take into account the reality and complexities of development (Gericke et al., 2014). Interestingly, even biology teachers may hold simplistic and inaccurate views of genetic determinism, as a study with teachers from 23 countries has shown (Castera and Clement, 2014). Thus, it should not be a big surprise to find out that people may finish high school possessing a naive deterministic view of genetics. This is, for example, the conclusion from a study of 500 high school students' essays, submitted for a national contest, in which genetic determinism was found to be a common conception (Mills Shaw et al., 2008). Research on secondary students' preconceptions about genetics also suggests that these are often difficult to overcome entirely during instruction (Duncan and Tseng, 2011; Tsui and Treagust, 2007). Regarding genetic determinism, secondary students often map the information included in DNA at higher organization levels than that of proteins. In other words, students map the molecular (genotypic) level directly onto higher levels (which may include that of the phenotype), overlooking the important role of proteins and of all other phenomena and mechanisms that result in the production of phenotypes (Duncan and Reiser, 2007).

It seems that the idea of genetic determinism is intuitive. A plausible explanation is that it stems from deep intuitions such as essentialism. Essentialism is the idea that entities have essences, i.e., underlying properties that are characteristic of them (see Wilkins, 2013 for a conceptual analysis). Psychological essentialism is the idea that the essences of organisms are fixed and unchanging, and as such they characterize organisms despite any superficial changes they may undergo (see Gelman, 2003 for a detailed discussion and Gelman, 2004 for a review). It has been suggested that essentialism has a strong impact on people's understanding of genetics. Genetic essentialism has been defined as our tendency "to infer a person's characteristics and behaviors from his or her perceived genetic makeup" (Dar-Nimrod and Heine, 2011; Nelkin and Lindee, 2004). In particular, it has been suggested that genetic essentialism may make people think of genetically influenced traits as unchangeable and determined, consider the relevant genes as being the fundamental cause of the respective character, view groups that share a genetic character as being homogeneous and discrete from other groups, and perceive characters as more natural if they are genetically determined (Dar-Nimrod and Heine, 2011). It has also been found that people tend to associate genes implicitly (and thus unconsciously) with immutable fate concepts (Gould and Heine, 2012). However, further research on this topic is required.

As readers of this book certainly know, genome-wide association studies have shown that the influence of single genetic factors on disease is small. Even for traits that run through families, the most probable candidate genetic factors account just for a small percentage of the overall inherited risk for a disease (Dermitzakis and Clark, 2009; Visscher et al., 2012). It seems that more than associations between DNA sequences and disease risk is required and so scientists' attention has turned toward understanding processes and mechanisms, such as the complex subcellular interactions among DNA, histones, and other molecules, making the DNA transcript and not the gene the focus of study (Stamatoyannopoulos, 2012). Given all this, and although the concept of "gene" is used continuously by researchers (who are well aware of what they are talking about), it might be better to refrain from referring to "genes" in the public discourse on genetics.

A viable alternative could be the concept of genetic material that is broader and can incorporate the complexities of developmental processes often overlooked in public discourse. The public often fails to realize that genes can do nothing outside their cellular contexts, and that the contribution of genes cannot strictly be distinguished from the contribution of their cellular and external environment. Such problems can be addressed if discourse is about genetic material, defined as "any nucleic acid with the propensity to be inherited and to interact with other cellular components as a source of sequence information, eventually affecting or being implicated in cellular processes with local or extended impact." This concept refers to particular macromolecules (DNA, but also RNA) that are related to the development and inheritance of traits, is not limited to particular functions because all functional parts of the genetic material may be implicated in various phenomena and phenotypes, and does not refer to contiguous DNA sequences because functional units may encompass different parts of the genome (Burian and Kampourakis, 2013).

27.2.2 "Crime Scene Investigation" Science: Portray an Authentic View of the Nature of Science

It seems that the mass media have an enormous influence on the public's conceptions about science as a process and about the kind of knowledge it produces, particularly when it comes to genetics. For instance, it has been found that even elementary schoolchildren consider TV shows to be their main source of information about genes and DNA (Donovan and Venville, 2014). This is especially important, given that popular TV series such as *CSI: Crime Scene Investigation* seem to advance an entirely unrealistic view of what can be achieved in science laboratories. In this case, such TV series impose a demand and a respect for forensic science that is not supported by its actual validity (Schweitzer and Saks, 2007). A detailed study of this show concluded that the dominant message is that DNA testing is common, swift, reliable, and instrumental in solving cases. It was also found that the show presents DNA processing as quick and easy, typically taking no more than a day to

complete, whereas this process may actually take days, weeks, or months (Ley et al., 2012). Most interestingly, it has also been found that TV shows that impose a view of science as certain enhance laypersons' interest in it, whereas those that impose a view of science as uncertain do not have such an effect (Retzbach et al., 2013). Consequently, if TV shows generally portray the conclusions of genetics testing as definitive and certain, and if this view is easily accommodated by the public, the latter will come to have unrealistic expectations from genetics and genomics and a wrong perception of the impact of the respective knowledge on societal issues.

It has also been found that films may enhance students' preconceptions about genetics, because the concepts conveyed about the structure and properties of the genetic material as well as its modification significantly overlap with students' preconceptions (Muela and Abril, 2014). Perhaps the most prominent example of how a film can serve as a reference point to discuss the ethical issues related to genetic technologies is *GATTACA*. This film presents an imaginative future in which, in contrast to the limited current use of human genetic technologies, parents are willing to improve the genetic makeup of their offspring. However, not everyone has access to the respective genetic technologies, and as a result there are two distinct groups of people: the dominant group of the genetically modified and the oppressed group of the genetically unmodified. Interestingly, the film does not misrepresent the potential of genetic technologies but rather warns that they may cause problems if a genetic determinist ideology that sees humans as the sum of their genes prevails in human societies. However, the film falls short in taking into account the problems of basing racial discrimination on differences in genes, because it presents one group of people as genetically inferior to another (Kirby, 2000, 2004).

Therefore, because TV series and films can influence public perception of and expectations about genetics, it is necessary to provide the public with a more accurate portrayal of how science is done and what knowledge it produces. Nowadays, one can easily find information about historical figures and the "discoveries" they made in textbooks, or the World Wide Web. However, often these accounts are distortions of how science has been or is currently being accomplished, reinforcing inaccurate conceptions. One of these is the implicit portrayal of science as a lonely endeavor in which a single person is presented as having been able to advance a whole scientific field. In some of these cases, such as Isaac Newton, scientific advancement was immediately achieved. In other cases, however, scientific advancement was not possible because the contemporaries of that individual were not able to understand the significance of his or her work because he or she was ahead of his or her time. Gregor Mendel is perhaps an exemplar of this stereotype. He is often presented as the lonely pioneer of genetics who died in oblivion and whose work was eventually recognized as laying the foundations of genetics. In particular, textbook accounts suggest that the study of genetics began with Mendel, that his studies were published as early as 1866 but that no one understood their importance, that his work was ignored for 34 years, and that scientists later discovered that it contained the entire basis of modern genetics, that he studied characters that could be classified on an either—or basis, that he observed that different alleles can produce different phenotypes, and even that he discovered that pairs separate independently during meiosis (Campanile et al., 2015).

However, if one carefully studies Mendel's paper, there are different conclusions to be made: that he intended to study hybridization and not heredity, that he studied how characters or traits were passed on from one generation to the next, that nowhere did he write about hereditary particles that determine these characters or traits, and that he was able to observe the segregation and independent assortment of characters or traits but he did not conclude anything about the segregation and the independent assortment of alleles (Kampourakis, 2015). The usual presentation of Mendel's life and work in textbooks also includes instances of general nature of science aspects that could be discussed in schools. In a study of seven widely used high school biology textbooks in terms of the aspects of the nature of scientific knowledge (NOSK) and of scientific inquiry (SI) that they presented explicitly or implicitly in the Mendelian genetics sections, many such aspects existed, but most were implicit. Overall, 237 instances of NOSK and 128 instances of SI aspects were counted in 140 textbook excerpts. Of these 365 instances, 362 were implicit and only 3 were explicit (all of which were SI aspects). The conclusion of this study was that the sections of Mendelian genetics might provide teachers with a multitude of opportunities to teach about NOSK and SI explicitly, yet these opportunities are likely to be missed because almost all of the relevant aspects are implicit (Campanile et al., 2015).

What does Mendel's story have to do with genetic testing? The point made here is that distortions of the nature of the science of genetics already exist in formal science education. Furthermore, opportunities are missed to teach about nature of science. But to understand the potential of science, it is necessary to convey a more authentic image of science in the teaching and public communication of genetics: that science is done by humans within communities, that it can be messy and that its conclusions are not always clear, that there can be conflicts of interest, and that the knowledge produced is not only tentative but also results in even more questions. This can be achieved if genetics education and outreach are explicit about aspects of nature of science, discuss the actual potential of contemporary genetics research, and explicitly distinguish it from future possibilities (Barnes and Dupré, 2008; Krimsky and Gruber, 2013; Kampourakis, 2013).

27.2.3 Lack of Public Awareness of Contemporary Genetics Research: Researchers Should Get Actively Involved in Science Communication

Current developments and advances in genetics research unavoidably lead to uncharted waters, because in the early stages of the application of new technologies there is no sufficient experience to foresee how it might be used or abused, which ethical and societal problems might arise, who might object to its use and why, and how it might be possible to deal with such problems. Therefore, as early as possible it is necessary to provide unbiased, well-informed, and well-argued estimates of potential questions and problems, as well as potential answers to these. Those developing the relevant technologies and conducting contemporary genetics research can provide these answers, because they are the ones who know their potential and their pitfalls. Therefore, it is necessary that they become actively involved in public discussions and also education and outreach, so that the public be well-informed about these. If this does not happen, there is always the possibility that the public will be misinformed and misguided by those who have particular interests to serve by promoting the use of the new technologies or by those who, for various reasons, are against any such use. The public needs to have a robust understanding and develop a critical stance before making any decisions. New research and technologies can be neither accepted nor rejected uncritically (Reydon et al., 2012).

Geneticists themselves would most likely agree that they should be engaged in genetics communication and outreach. However, as a study involving 20 academic geneticists in the United States showed, there are particular barriers to their participation in these kinds of activities, including: (1) lack of time to engage in activities besides those required to earn tenure; (2) receipt of no credit toward tenure for participating in public engagement activities; (3) fear of professional stigmatization by being too publicly visible; (4) lack of awareness of opportunities for public engagement and participation in policy formation; and (5) lack of necessary communications skills (Mathews et al., 2005). Although these views were expressed by a small number of geneticists, and thus it is difficult to make generalizations, it seems likely that many geneticists would agree that several of these barriers are real. Time and recognition of such activities are probably major issues for many researchers. Therefore, perhaps universities should reconsider their criteria for tenured professorships and evaluate candidates in terms not only of their publishing record and research agendas, but also of their participation in outreach activities and science communication.

Another major issue is how researchers in genetics communicate their research and its results and conclusions to journalists, who in turn communicate these to the public. Condit (2007) analyzed this issue and made some concrete suggestions that might facilitate a more accurate communication between researchers and journalists. The first suggestion is that instead of just chatting with the journalist over the phone or during an interview, researchers should prepare the message they want to convey by writing the key points in lay language and testing its readability. The second suggestion is for researchers to try to avoid hype that is enhanced by their enthusiasm for their own work and the demands for funding. Thus, instead of making ambitious predictions for future applications, they should describe the technical and social roadblocks and explain why current findings could be overturned. A third suggestion is to avoid gene-centered frames, such as "gene for" and "genes win," and in general genetic determinism that seems so intuitive to nongeneticists (see also the earlier respective section). Because two organisms with the same genes may have different phenotypes in different environments, geneticists should make it clear that it is the interaction of genes and environment that affects outcomes. For this purpose, Condit suggests that it is also better to talk about "alleles" or "versions of a gene" rather than "the gene," because people tend to think that the gene is something that someone either has or has not rather than something that may have different versions. Finally, the fourth suggestion is to avoid labels that enhance discrimination among different human groups or populations, such as intuitive notions of "race," or referring to people carrying an allele that is related to a specific disease. Instead, geneticists should try to characterize human groups or populations carefully and precisely.

One important issue besides the need to present the actual potential of genetic technologies accurately is the accurate presentation of concepts. One of the most distorted concepts in the news is that of "heritability." This concept is commonly defined as the extent to which genes and the environment affect the development of a trait. Thus, for instance, when one reads that the heritability of a given trait is 40%, this is perceived as a description of the fact that the influence of genes is 40% and the influence of the environment is 60%: in other words, that the particular trait is 40% genetic and 60% environmental. However, this is far from true because one cannot separate the two influences. To use a metaphor, one cannot say whether the performance of a drummer is more due to his or her talent or to the quality of the instruments he or she is playing. Such a point would make sense only if one compared: (1) two drummers playing the same or different instruments, in which case any difference would be 100% due to the drummers; (2) one drummer playing different instruments, in which case any difference would be 100% due to the instruments; or (3) two drummers playing different instruments, in which case the difference would be somewhere between (1) and (2) (Keller, 2010, pp. 34–36). Therefore, to state that the heritability of a trait is 40% actually reflects the extent to which variation in a trait across a population can be accounted for by variation in genes across that

population. This has nothing to do with how much "genetic" or "inherited" a trait is or how strong the effect of genes over environment is during development. To give one example, the fact that humans normally have five digits definitely depends on genetic factors rather than on environmental ones, and so humans everywhere in the world are normally born with five digits. Yet, because these genetic factors do not vary significantly in human populations and because any difference in the number of digits would most likely be the result of environmental influences before or after birth, the heritability of this trait could be close to zero (Moore, 2013).

In November 2011, the National Human Genome Research Institute held a meeting to address the question of how to achieve genomic literacy for the general public. The workshop participants agreed on four priorities: (1) Conduct research to define particular genomic literacy competencies that focus not on what the public should know about genetics and genomics but rather what they should be able to do given their knowledge, skills, and attitudes relevant to genetics and genomics; (2) conduct systematic research on which the best education practices are for raising the genomic literacy and make this research available to those interested, such as health care providers, researchers, and educators; (3) produce and make available "genomic literacy assessment" tools, which are necessary for systematically documenting the public understanding of genetics and genomics over time; and (4) establish a genomics education national conference to promote the interaction and exchange of ideas, research, and resources among all those interested in genetics literacy (Hurle et al., 2013). Initiatives such as these are likely to lead to systematic ways of science communication and outreach, which might significantly contribute to enhancing the genetics literacy of the public.

27.3 CONCLUSIONS AND SUGGESTIONS

The aims of this chapter were to explain why the public understanding of genetics is a necessary condition for genetics literacy, provide evidence of the low public understanding of genetics, and present some crucial obstacles that should be overcome to achieve literacy in genetics: (1) the prevalence of genetic determinism in genetics education that enhances the notion that there exist "genes for traits"; (2) the poor understanding of the nature of the science of genetics that creates unrealistic expectations by overlooking the actual potential; and (3) the limited access that the public has to contemporary genetics research owing to the limited involvement of researchers in science communication. To overcome these obstacles, we need to transform genetics education and outreach to present a more accurate portrayal of the role of genes and of the nature of the science of genetics. These can be achieved by revising curricula and textbooks, appropriately training teachers, and involving geneticists in education and outreach. Education and public communication of genetics should aim to provide an authentic picture with all of the conceptual and practical complexities of genetics.

There are particular considerations to make for this purpose. First, the public needs to understand that the meaning of the concept of gene has changed since its inception in 1909. An initially "empty" or referentially indefinite concept of gene came to have two distinct meanings: that of a hypothetical inherited factor that related to changes in characters, and that of a DNA sequence that encoded the information for a character. These two gene concepts did not converge into the same pieces of DNA. Second, the public needs to understand that reference to single genes cannot explain the variation observed not only for complex characters and disease but also for simple monogenic ones. Single genes alone do not produce characters or disease, but differences in genes can cause differences in traits and disease. Causation and associations are also different. Third, the public needs to understand that genes "function" in the context of developmental processes only, and that there important phenomena of nongenetic inheritance exist, i.e., beyond DNA and its sequence. They also need to understand better the interplay between our genetic makeup ("nature") and the environmental influences on them during development and beyond ("nurture"). It is with such a background that the public could better understand important socio-scientific issues and debates relevant to genetics: personalized medicine, forensic investigations, ancestry and race, behavior, and genetic engineering. Making sense of genes should become a central objective of genetics literacy (Kampourakis, 2017).

REFERENCES

Barnes, B., Dupré, J., 2008. Genomes and What to Make of Them. University of Chicago Press, Chicago.

Beckwith, J., 2002. Genetics in society: society in genetics. In: Alper, J., Ard, C., Asch, A., Beckwith, J., Conrad, P., Geller, L. (Eds.), The Double-Edged Helix: Social Implication of Genetics in a Diverse Society. Johns Hopkins University Press, Baltimore, pp. 39−57.

Borzekowski, D.L.G., Guan, Y., Smith, K.C., Erby, L.H., Roter, D.L., 2013. The Angelina effect: immediate reach, grasp, and impact of going public. Genet. Med. 16, 516−521.

Burian, R.M., Kampourakis, K., 2013. Against "genes for": could an inclusive concept of genetic material effectively replace gene concepts? In: Kampourakis, K. (Ed.), The Philosophy of Biology: A Companion for Educators. Springer, Dordrecht, pp. 597−628.

Campanile, M.F., Lederman, N.G., Kampourakis, K., 2015. Mendelian genetics as a platform for teaching about nature of science and scientific inquiry: the value of textbooks. Sci. & Educ. 24 (1−2), 205−225.

Castera, J., Clement, P., 2014. Teachers' conceptions about genetic determinism of human behaviour: a survey in 23 countries. Sci. & Educ. 23 (2), 417–443.

Condit, C.M., 2010a. Public attitudes and beliefs about genetics. Annu. Rev. Genomics Hum. Genet. 11, 339–359.

Condit, C.M., 2010b. Public understandings of genetics and health. Clin. Genet. 77, 1–9.

Condit, C.M., Ofulue, N., Sheedy, K.M., 1998. Determinism and mass-media portrayals of genetics. Am. J. Hum. Genet. 62, 979–984.

Condit, C.M., 2007. How geneticists can help reporters to get their story right. Nat. Rev. Genet. 8, 815–820.

Dar-Nimrod, I., Heine, S.J., 2011. Genetic essentialism: on the deceptive determinism of DNA. Psychol. Bull. 137 (5), 800–818.

Dermitzakis, E.T., Clark, A.G., 2009. Life after GWA studies. Science 326 (5950), 239–240.

Donovan, J., Venville, G., 2014. Blood and bones: the influence of the mass media on Australian primary school children's understandings of genes and DNA. Sci. & Educ. 23 (2), 325–360.

Duncan, R.G., Tseng, K.A., 2011. Designing project-based instruction to foster generative and mechanistic understandings in genetics. Sci. Educ. 95 (1), 21–56.

Duncan, R.G., Reiser, B.J., 2007. Reasoning across ontologically distinct levels: students' understandings of molecular genetics. J. Res. Sci. Teach. 938–959.

Eurydice Network, 2011. Science Education in Europe: National Policies, Practices and Research. Education. Audiovisual and Culture Executive Agency, Brussels.

Gelman, S.A., 2004. Psychological essentialism in children. Trends Cogn. Sci. 8 (9), 404–409.

Gelman, S.A., 2003. The Essential Child: Origins of Essentialism in Everyday Thought. Oxford University Press, Oxford.

Gericke, N., Hagberg, M., Carvalho Santos, V., Joaquim, L.M., El-Hani, C., 2014. Conceptual variation or incoherence? Textbook discourse on genes in six countries. Sci. & Educ. 23 (2), 381–416.

Goldsmith, L., Jackson, L., O'Connor, A., Skirton, H., 2012. Direct-to-consumer genomic testing: systematic review of the literature on user perspectives. Eur. J. Hum. Genet. 8, 811–816.

Gould, W.A., Heine, S.J., 2012. Implicit essentialism: genetic concepts are implicitly associated with fate concepts. PLoS One 7 (6), e38176. http://dx.doi.org/10.1371/journal.pone.0038176.

Guttmacher, A.E., Porteous, M.E., McInerney, J.D., 2007. Educating health-care professionals about genetics and genomics. Nat. Rev. Genet. 8 (2), 151–157.

Haga, S., Burke, W., Ginsburg, G., Mills, R., Agans, R., 2012a. Primary care physicians' knowledge of and experience with pharmacogenomics testing. Clin. Genet. 82 (4), 388–394.

Haga, S.B., O'Daniel, J.M., Tindall, G.M., Mills, R., Lipkus, I.M., Agans, R., 2012b. Survey of genetic counselors and clinical geneticists' use and attitudes toward pharmacogenetic testing. Clin. Genet. 82, 115–120.

Hurle, B., Citrin, T., Jenkins, J.F., Kaphingst, K.A., Lamb, N., Roseman, J.E., Bonham, V.L., 2013. What does it mean to be genomically literate?: National Human Genome Research Institute Meeting Report. Genet. Med. 15 (8), 658–663.

Jiménez-Aleixandre, M.P., 2014. Determinism and underdetermination in genetics: implications for students' engagement in argumentation and epistemic practices. Sci. & Educ. 23 (2), 465–484.

Kampourakis, K., 2013. Mendel and the path to genetics: portraying science as a social process. Sci. & Educ. 22 (2), 293–324.

Kampourakis, K., 2015. Myth 16: that Mendel was a pioneer of genetics, being ahead of his time. In: Numbers, R.L., Kampourakis, K. (Eds.), Newton's Apple and Other Myths about Science. Harvard University Press, Cambridge MA, pp. 129–138.

Kampourakis, K., 2017. Making Sense of Genes. Cambridge University Press, Cambridge.

Kampourakis, K., Vayena, E., Mitropoulou, C., van Schaik, R.H., Cooper, D.N., Borg, J., Patrinos, G.P., 2014. Next-generation pharmacogenomics and society: key challenges ahead. EMBO Rep. 15 (5), 472–476.

Keller, E.F., 2010. The Mirage of a Space between Nature and Nurture. Duke University Press.

Kirby, D.A., 2000. The new eugenics in cinema: genetic determinism and gene therapy in GATTACA. Sci. Fiction Stud. 27 (2), 193–215.

Kirby, D.A., 2004. Extrapolating race in GATTACA: genetic passing, identity, and the science of race. Lit. Med. 23 (1), 184–200.

Kitcher, P., 2003. Battling the undead: how (and how not) to resist genetic determinism. In: Kitcher, P. (Ed.), In Mendel's Mirror: Philosophical Reflections on Biology. Oxford University Press, Oxford, pp. 283–300.

Krimsky, S., Gruber, J. (Eds.), 2013. Genetic Explanations: Sense and Nonsense. Harvard University Press, Cambridge MA.

Lachance, C.R., Erby, L.A., Ford, B.M., Allen, V.C., Kaphingst, K.A., 2010. Informational content, literacy demands, and usability of websites offering health-related genetic tests directly to consumers. Genet. Med. 12 (5), 304–312.

Leighton, J.W., Valverde, K., Bernhardt, B.A., 2012. The general public's understanding and perception of direct-to-consumer genetic test results. Public Health Genomics 15 (1), 11–21.

Lewis, J., Leach, J., 2006. Discussion of socio-scientific issues: the role of science knowledge. Int. J. Sci. Educ. 28 (11), 1267–1287.

Ley, B.L., Jankowski, N., Brewer, P.R., 2012. Investigating *CSI*: portrayals of DNA testing on a forensic crime show and their potential effects. Public Underst. Sci. 21 (1), 51–67.

Mathews, D.J.H., Kalfoglou, A., Hudson, K., 2005. Geneticists' views on science policy formation and public outreach. Am. J. Med. Genet. 137A, 161–169.

Mills Shaw, K.R., Van Horne, K., Zhang, H., Boughman, J., 2008. Essay contest reveals misconceptions of high school students in genetics content. Genetics 178 (3), 1157–1168.

Moore, D.S., 2008. Espousing interactions and fielding reactions: addressing laypeople's beliefs about genetic determinism. Philos. Psychol. 21 (3), 331–348.

Moore, D.S., 2013. Current thinking about nature and nurture. In: Kampourakis, K. (Ed.), The Philosophy of Biology: A Companion for Educators. Springer, Dordrecht, pp. 629–652.

Muela, F.J., Abril, A.M., 2014. Genetics and cinema: personal misconceptions that constitute obstacles to learning. Int. J. Sci. Educ. Part B Commun. Public Engagem. 4 (3), 260–280.

National Research Council Committee on Conceptual Framework for the New K-12 Science Education Standards, 2012. A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. National Academies Press, Washington.

Nelkin, D., Lindee, S.M., 2004. The DNA Mystique: The Gene as a Cultural Icon. University of Michigan Press, Ann Arbor.

Retzbach, J., Retzbach, A., Maier, M., Otto, L., Rahnke, M., 2013. Effects of repeated exposure to science TV shows on beliefs about scientific evidence and interest in science. J. Media Psychol. 25 (1), 3−13.

Reydon, T.R., Kampourakis, K., Patrinos, G.P., 2012. Genetics, genomics and society: the responsibilities of scientists for science communication and education. Pers. Med. 9 (6), 633−643.

Roberts, D.A., 2007. Scientific literacy/science literacy. In: Abell, S.K., Lederman, N.G. (Eds.), Handbook of Research on Science Education. Erlbaum, Mahwah, NJ, pp. 729−780.

Sadler, T.D., Fowler, S.R., 2006. A threshold model of content knowledge transfer for socioscientific argumentation. Sci. Educ. 90 (6), 986−1004.

Sadler, T.D., Zeidler, D.L., 2005. The significance of content knowledge for informal reasoning regarding socioscientific issues: applying genetics knowledge to genetic engineering issues. Sci. Educ. 89 (1), 71−93.

Schweitzer, N.J., Saks, M.J., 2007. The *CSI* Effect: popular fiction about forensic science affects the public's expectations about real forensic science. Jurimetrics 47, 357−364.

Stamatoyannopoulos, J.A., 2012. What does our genome encode? Genome Res. 22, 1602−1611.

Stanek, E.J., Sanders, C.L., Johansen Taber, K.A., Khalid, M., Patel, A., Verbrugge, R.R., Agatep, B.C., Aubert, R.E., Epstein, R.S., Frueh, F.W., 2012. Adoption of pharmacogenomic testing by US physicians: results of a nationwide survey. Clin. Pharmacol. Ther. 91 (3), 450−458.

Thompson, W.C., 2013. Forensic DNA evidence: the myth of infallibility. In: Krimsky, S., Gruber, J. (Eds.), Genetic Explanations: Sense and Nonsense. Harvard University Press, Cambridge MA, pp. 227−255.

Tsui, C.Y., Treagust, D.F., 2007. Understanding genetics: analysis of secondary students' conceptual status. J. Res. Sci. Teach. 44 (2), 205−235.

Visscher, P.M., Brown, M.A., McCarthy, M.I., Yang, J., 2012. Five years of GWAS discovery. Am. J. Hum. Genet. 90 (1), 7−24.

Wilkins, J.S., 2013. Essentialism in biology. In: Kampourakis, K. (Ed.), The Philosophy of Biology: A Companion for Educators. Springer, Dordrecht, pp. 395−419.

Chapter 28

Safety and the Biorepository

D.S. Atherton[1], J. Fredenburgh[2], K.C. Sexton[1], W.C. Bell[1] and W.E. Grizzle[1]
[1]University of Alabama at Birmingham, Birmingham, AL, United States; [2]Scigen Scientific Inc., Gardena, CA, United States

28.1 INTRODUCTION

In the operation of a biorepository, a number of potential safety hazards must be considered. These range from chemical and biological hazards associated with the routine use of reagents and human and animal tissue to fire, electrical, and physical hazards associated with laboratory infrastructure. An individual biorepository's set of safety issues depends on the specific roles and goals of that biorepository. For example, a biorepository that specifically collects tissues from patients with blood-borne pathogens would be much more concerned with associated biological hazards than would a biorepository that collected only tissue from individuals who were not at risk, although reasonable precautions should be in place in both types of systems to minimize the exposure of laboratory personnel to pathogens.

Similarly, a laboratory that extracts and/or studies DNA/RNA may have unique safety issues, especially if recombinant DNA and/or synthetic nucleic acids are used. The safety precautions associated with recombinant and synthetic DNA/RNA are complex, particularly when infective organisms are the source of the nucleic acids or the DNA could code for toxins. Safety issues associated with any work with recombinant DNA/RNA are postulated to have special hazards and are beyond the scope of this chapter. The reader is referred to the National Institutes of Health Guidelines for Research Involving Recombinant or Synthetic Nucleic Acid Molecules (NIH Guidelines for Research, 2013).

The routine extraction of DNA and RNA from human and animal tissues has safety hazards that fall into the categories subsequently discussed in this chapter, including safety issues associated primarily with biohazards, chemical hazards, and chemical and biohazard disposal. Using tissue for the extraction of DNA/RNA involves the same biohazards associated with handling any human tissue (i.e., universal precautions). Typically, nucleic acids are extracted from frozen, fresh nonfrozen, or paraffin-embedded tissues using commercial kits that usually contain some of the chemicals necessary for extraction. These "kits" always contain information on safety as well as safety information on any chemical reagents in the kit. Extracting DNA and RNA from paraffin-embedded tissues requires removal of the paraffin using xylene and rehydrating the tissues using graded ethanols. Xylene and the graded ethanols are not provided in typical commercial kits used in extracting DNA/RNA from paraffin-embedded tissues and have to be obtained separately as they would if a kit were not used for extraction. Similarly, if a commercial kit is not used in the extraction process, all chemicals must be obtained separately. These chemicals will have safety data sheets (SDS) provided by the manufacturer or supplier.

Several countries have organizations that regulate laboratories and offer general guidelines intended to protect laboratory employees. In the United States, the Occupational Safety and Health Administration (OSHA) issues regulations, including those designed to limit hazardous chemical exposure, to ensure employees have access to hazard information and address the use of personal protective equipment (PPE) in the workplace. In addition, biorepositories in the United States may adhere to fire and electrical safety guidelines established by the National Fire Protection Association (NFPA). Biorepositories in countries in the European Union (EU) follow similar guidelines with their own respective regulatory agencies. In addition, many regional (e.g., state or province) and local (e.g., city) jurisdictions provide their own safety regulations, especially in the areas of physical, fire, and electrical safety.

The goal of this chapter is to aid biorepositories in developing and improving their safety program based on the experience of the Tissue Collection and Banking Facility (TCBF) at the University of Alabama at Birmingham

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00028-6
Copyright © 2017 Elsevier Ltd. All rights reserved.

(UAB), and of the authors. This chapter *does not contain all of the information necessary to ensure the safety of all biorepository personnel or to ensure a laboratory can meet all regulatory, certification, and/or accreditation standards in safety*. However, it provides a foundation upon which a biorepository can develop and maintain a functional safety program that results in minimal hazardous exposure and workplace injuries to laboratory employees. This chapter only provides a foundation, and maintenance of a functional safety program requires constant attention to new issues related to laboratory safety that frequently develop. For example, in the United States, chemical safety regulations are changing to be more consistent with the Globally Harmonized System (GHS) of the Classification and Labeling of Chemicals of the World Health Organization (WHO) of the United Nations (UN). Literature pertinent to this transition includes Hazard Communication Standards (Occupational Safety; Occupational Safety and Health), *A Guide to the GHS* (Safety & Health Administration), and websites listed under Chemical Safety in Table 28.1.

28.2 UNDERSTANDING REGULATORY AND OTHER SAFETY ISSUES RELEVANT TO BIOREPOSITORIES

Regulatory agencies appropriately place a high priority on the general health and safety of laboratory personnel. Abundant guidelines are available to which biorepositories can refer to assist in developing and maintaining their safety programs. Table 28.1 lists several Web-based resources that describe several facets of safety regulations in the United States and EU, and other countries. Several articles are also available (Grizzle and Polt, 1988; Beekmann and Doebbeling, 1997; Cardo and Bell, 1997; Grizzle and Fredenburgh, 2001; Grizzle et al., 2010), in addition to three versions of the Best Practices of the International Society of Biological Environmental Repositories (Astrin et al., 2012; Pitt et al., 2008; Aamodt et al., 2005) and various books (Table 28.2). All of these resources can help develop a biorepository's safety program, whether the biorepository falls under an academic setting, commercial setting, or government setting. Notably, biorepositories associated with a hospital's anatomic and clinical pathology (AP/CP) department will also need to consider safety regulations pertaining to the function of the AP/CP department, many of which are similar to safety regulations specifically important to biorepositories.

A major goal of any laboratory or biorepository is to minimize injury to laboratory personnel and damage to the biorepository. As mentioned, often a biorepository's safety program is subject to regulations of a larger laboratory like an AP/CP department. However, this is not always the case, and biorepositories have varying levels of autonomy for the development and maintenance of their safety programs. In all cases, safety regulations ought to be transparent, and administrators and supervisors should be cognizant of safety regulations and the intended safety program that is in place. Developing and maintaining a well-run safety program require communication, and one of the most important interactions is between laboratory supervisors and technicians or other laboratory personnel performing technical duties. Laboratory technicians are often the individuals handling hazardous materials and potentially pathogenic tissues most frequently, which makes their understanding of safety regulations and practices important not only from the standpoint of their own protection but also because they often first recognize potential deficiencies in safety practices and can provide valuable input based on their experiences in the individual areas in which they work.

A safety plan also needs to incorporate engineering issues such as correcting problems with ventilation, drainage, infrastructure, and space and should also address the appropriate use of safety equipment. One method to ensure all laboratory personnel are familiar with the appropriate use of safety equipment is through education, which is one of the most important parts of any laboratory's safety program. Essentially all laboratory personnel involved in the biorepository need systematic training on necessary elements of the safety program, including but not limited to safety regarding blood-borne pathogens, chemical hazards, formaldehyde, and numerous other general safety issues.

28.3 INDIVIDUALS INVOLVED IN OVERSIGHT OF A BIOREPOSITORY

The chief executive officer (CEO) of an organization or hospital usually appoints a safety committee whose general function is to oversee safety within the organization, but the CEO is ultimately responsible for the safety of the organization's employees. Because different laboratories within organizations have different roles, the safety requirements of each laboratory are often unique. Sometimes, safety committees can be even further subdivided into subcommittees, with each subcommittee given specific oversight over certain aspects of safety or oversight over a specific laboratory or group of laboratories. The safety committee is also responsible for periodically reviewing a laboratory's safety program, recommending modifications as needed, and reviewing incidents and how to handle and prevent them in the future.

The safety committee and CEO of an organization can also appoint a safety officer (SO), an individual member of a safety committee who is specifically tasked with certain aspects of safety oversight. For example, the SO may be

TABLE 28.1 Resources on Safety on the Internet

Website	Organization	Topics
Safety in General		
http://www.osha.gov	Occupational Safety and Health Administration (OSHA), Department of Labor, United States	Current developmental and operational regulations; technical information; prevention information; training information; links to other sites
http://www.ccohs.ca/oshanswers	Canadian Center for Occupational Health and Safety	Answers on frequently asked questions on safety
https://www.osha.europa.eu/en/legislation/index_html	Occupational and Health Administration, European Union	General safety, European Union
https://www.osha.europa	Occupational and Health Administration, European Union	General safety, European Union
http://www.rmlibrary.com/db/lawosha.htm	Libraries and Directories: Risk Management and Insurance Safety	Occupational safety laws of all 50 states of United States
http://www.cap.org	College of American Pathologists	General and technical information; laboratory management; laboratory safety
http://www.clsi.org/	Clinical and Laboratory Standards Institute	General and technical information; forms; safety; links
http://www.lbl.gov/ehs/pub3000	Lawrence Berkeley National Laboratory Health and Safety	Health and safety manual
http://www.healthsystem.virginia.edu/internet/epinet/	University of Virginia, International Health Care Worker Safety Center	Surveillance data
http://www.healthsystem.virginia.edu/internet/epinet/about_epinet.cfm	Exposure Prevention Information Network (EpiNet)	Surveillance data
https://www.osha.gov/dsg/hazcom/HCSFinalRegTxt.html	OSHA, United States	Safety aids
Biological Safety		
http://www.cdc.gov	Centers for Disease Control and Prevention, Atlanta, GA	Surveillance data; prevention information; technical information; biohazards; links
http://www.absa.org	American Biological Safety Association	Technical information
http://www.cdc.gov/ncezid/	National Center for Emerging and Zoonotic Infectious Diseases	Combines National Center for Infectious Disease and National Center for Zoonotic, Vector-Borne and Enteric Diseases
http://www.cdc.gov/nchhstp/	National Center for HIV/AIDS, Viral Hepatitis, STD, and Tuberculosis (TB) Prevention	Biological information on HIV/AIDS, viral hepatitis, STD, and TB prevention
http://www.cdc.gov/ncird/	National Center for Immunization and Respiratory Diseases	Biological information on immunization and respiratory diseases
http://www.fda.gov/cber	Food and Drug Administration, Center for Biologics Evaluation and Research	Information on recalls, withdrawals, and safety issues concerning biologics

Continued

TABLE 28.1 Resources on Safety on the Internet—cont'd

Website	Organization	Topics
http://npic.orst.edu/	National Pesticide Information Center	Technical information on disinfectants; links to other sites
http://www.epa.gov/	Environmental Protection Agency	
http://www.defra.gov.uk/.	UK Department for Environment, Food, and Rural Affairs	Surveillance data on bovine spongiform encephalopathy in Europe; technical information on prions
http://www.cjd.ed.ac.uk	National Creutzfeldt–Jakob Disease (CJD) Surveillance Unit	Surveillance data on CJD; technical information on prions; links to other sites
http://www.who.int/csr/resources/publications/biosafety/WHO_CDS_CSR_LYO_2004_11/en/	World Health Organization (WHO) Laboratory Biosafety Manual	Biosafety information
Chemical Safety		
https://www.osha.gov/dsg/hazcom/index/html	OSHA, United States	Frequently asked chemical safety questions
http://www.ccohs.ca/oshanswers/chemicals/ghs.html	Canadian Center for Occupational Health and Safety	Chemical hazard questions
http://www.unece.org/trans/danger/publi/ghs/ghs_welcome_e.html	United Nations Economic Commission for Europe	Chemical safety
http://www.ghslegislation.com/asia-pacific/ghs-implementation-in-asia-pacific-countries-2/	Globally Harmonized System Legislation	Chemical regulation Asia–Pacific
http://www.cdc.gov/niosh/database.html	National Institute for Occupational Safety and Health (NIOSH)	Databases and information resource links and publications
http://www.ilo.org/public/english/protection/safework/cis/products/icsc/dtasht/index.htm	International Occupational Safety and Health Information Center	Chemical database; international Chemical safety cards
http://www.response.restoration.noaa.gov/chemaids/react.html	Office of Response and Restoration	Chemical reactivity worksheet; chemical database of reactivity of substances or mixture of substances
http://www.cdc.gov/niosh/chem-inx.html	Master Index of Occupational Health Guidelines for Chemical Hazards (NIOSH)	Guidelines for hazards of specific chemicals
http://www.who.int/ipcs/en/	WHO International Program on Chemical Safety	Chemical safety information
http://www.who.int/ipcs/methods/harmonization/areas/ra_toolkit/en/	WHO Human Health Risk Assessment Toolkit: Chemical Hazards	Chemical safety information
Fire Safety		
http://www.lbl.gov/ehs/pub3000/ch12.html	Lawrence Berkeley National Laboratory Health and Safety	Fire prevention and protection program
http://www.stonybrook.edu/ehs/fire	Stony Brook University Environmental Health and Safety	Laboratory fire safety hazard assessment and work practices
http://www.nfpa.org	National Fire Protection Association	Codes and standards, safety information and training in fire prevention

Electrical Safety		
http://www.lbl.gov/ehs/pub3000/CH14.html	Lawrence Berkeley National Laboratory Health and Safety	Electrical equipment safety program
http://web.princeton.edu/sites/ehs/labguide/sec_5.htm	Princeton University	Laboratory electrical safety program
http://www.ehs.uconn.edu/occ/elec.pdf	University of Connecticut Environmental Health and Safety	Electrical safety in laboratory
http://www.nfpa.org	National Fire Protection Association	Electrical codes related to fire protection
Physical Safety		
http://www.nfpa.org	National Fire Protection Association	Standards for storage of cryogenic fluids (NFPA.55 and NFPA.99)

TABLE 28.2 Books on Safety
Biosafety in the Laboratory, 1989. Prudent Practices for the Handling and Disposal of Infectious Materials. National Academies Press, Washington, DC.
Block, S.S., 2001. Disinfection, Sterilization, and Preservation, fifth ed. Lea and Febiger, Philadelphia, PA.
Bloom, B.R., 1994. Tuberculosis: Pathogenesis, Protection, and Control. American Society for Microbiology Press, Washington, DC, pp. 85–110.
Centers for Disease Control and Prevention/National Institutes of Health, 1999. Biosafety in Microbiological and Biomedical Laboratories, fourth ed. US Government Printing Office: US Department of Health and Human Services, Public Health Service, CDC and NIH, Washington, DC http://www.cdc.gov/biosafety/publications/bmb15/BMBL.pdf.
Fleming, D.O., Richardson, J.H., Tulis, J.H., Wesley, D., 1995. Laboratory Safety: Principles and Practices, second ed. American Society for Microbiology Press, Washington, DC.
Fredenburgh, J.L., Grizzle, W.E., 1993. Safety and Compliance in the Histology Laboratory: Biohazards to Toxic Chemicals (available only at workshops).
Furr, A.K., 2000. CRC Handbook of Laboratory Safety, fifth ed. CRC Press, Boca Raton, FL.
Gile, T.J., 2010. Complete Guide to Laboratory Safety, third ed. HCPro Inc., Marblehead, MA.
The globally harmonized system of classification and labeling of chemicals (GHS): Copyright United Nations, 2013, Rev. 5 http://www.unece.org/trans/danger/publi/ghs/ghs_rev05/05files_e.html.
Heinsohn, P.A., Jacobs, R.R, Concoby, B.A., 1996. AIHA Biosafety Reference Manual, second ed. American Industrial Hygiene Association, Fairfax, VA.
Kent, P.T., Kubica, G.P., 1985. Public Health Mycobacteriology. A Guide for the Level III Laboratory. US Department of Health and Human Services, Public Health Service, CDC, Atlanta, GA.
Kubica, G.P., Dye, W.E., 1967. Laboratory Methods for Clinical and Public Health Mycobacteriology. Public Health Service Publication Number 1547. US Department of Health, Education, and Welfare. United States Government Printing Office, Washington, DC.
Lieberman, D.F., 1995. Biohazards Management Handbook. Marcel Dekker, New York.
Miller, B.M., 1986 Laboratory Safety Principles and Practices. American Society for Microbiology, Washington, DC.
OSHA QuickTakes, 2012. A Guide to the Globally Harmonized System of Classification and Labeling of Chemicals. United States Department of Labor, Washington, DC https://www.osha.gov/dsg/hazcom/ghs.html.
Padhye, A.A., Bennett, J.E., McGinnis, M.R., et al. 1998. Biosafety considerations in handling medically important fungi. Med. Mycol. 36, 258–265.
Preventing Occupational Disease and Injury, 1991. American Public Health Association, Washington, DC.
Richmond, J.Y., Knudsen, R.C., Good, R.C., 1996. Biosafety in the clinical mycobacteriology laboratory. Clin. Lab. Med. 16, 527–550.
Richmond, J.Y., 1997. Designing a Modern Microbiological/Biomedical Laboratory: Lab Design and Process and Technology. American Biological Safety Association, Mundelein, IL.
Richmond, J.Y., 2000a. Anthology of Biosafety I: Perspectives on Laboratory Design. American Biological Safety Association, Mundelein, IL.
Richmond, J.Y., 2000b. Anthology of Biosafety II: Facility Design Considerations. American Biological Safety Association, Mundelein, IL.
Richmond, J.Y., 2000c. Anthology of Biosafety III: Application of Principles. American Biological Safety Association, Mundelein, IL.
Sewell, D.L., 1995. Laboratory-associated infections and biosafety. Clin. Microbiol. Rev. 8, 389–405.
Wald, P.H., Stave, G., 2001. Physical and Biological Hazards of the Workplace, second ed. Van Nostrand Reinhold, New York.

given personal oversight regarding a laboratory's safety plan, reviewing its effectiveness and recommending modifications. This includes monitoring the proper use of equipment, monitoring the environment, and ensuring employees' compliance with recommended safety practices and preventative health care (e.g., a requirement for hepatitis B vaccination of employees).

28.4 SAFETY TRAINING/EMPLOYEE EDUCATION IN A BIOREPOSITORY

Proper training in each specific safety area (e.g., biohazard safety such as education in blood-borne pathogens, or physical safety) generally follows a standardized approach. Laboratory personnel should be appropriately and thoroughly trained before they begin work. As employees shift duties or take on new tasks, they should undergo safety training relevant to those new tasks, ideally before they begin working on the new tasks.

Adequate training of personnel requires knowledgeable trainers and requires educational programs that are appropriate for the level, education, and language of the employee. The training should be commensurate with the safety issues for which that specific employee requires education (e.g., employees of biorepositories that do not use or store chemicals may not require education training in those biohazards). In addition, some individuals employed by the laboratory (i.e., janitors or secretaries) may not require education regarding all aspects of the laboratory, such as biohazardous materials or glassware safety. However, these individuals could inadvertently be exposed to some biohazardous materials via contaminated surfaces, paper, or computer equipment, etc. Thus, general awareness of safety issues beyond that which is necessarily trained may be beneficial not only for laboratory technicians but for other individuals as well, and all individuals should be educated in pertinent safety issues accordingly.

Because regulations change periodically, employees may require frequent updating regarding various aspects of training, and many elements of training are to be documented annually. The United States requires records of training be kept for at least 3 years. Employees' training and proper documentation of that training by the safety committee or SO can be a part of the quality management system of a biorepository's safety program (Grizzle et al., 2015).

28.5 BIOREPOSITORY SAFETY AREAS

Major areas of safety that should be addressed by biorepositories in their safety programs include biohazards, chemical hazards, physical security and safety, fire safety, and electrical safety.

28.5.1 Biohazards

Biorepositories are typically responsible for the collection, processing, storage, and distribution of human and/or animal tissue. Both human and animal tissue can potentially contain pathogens that can infect humans. Examples of blood-borne pathogens include HIV and selected members of the family of hepatitis viruses (e.g., hepatitis C virus). Thus, both human and animal tissue should be treated with universal precautions to prevent transmission of organisms contained within tissue to laboratory personnel; as such, a preventative plan should be in place to minimize exposure to highly infectious agents, especially blood-borne pathogens. Of note, organisms in some animal tissues as well as animal models of infectious diseases of humans can be transmitted to humans. Typically, fresh tissues contain the most viable infectious organisms, and the risk of transmission tends to decrease as tissue is processed (Fig. 28.1). Formalin fixation is a particularly important step in tissue processing, because formalin solution causes a large decline in the infectivity of bacteria and viruses within tissue. Most formalin-fixed tissues are eventually embedded in paraffin blocks, and biorepositories ought to be aware that even paraffin-embedded tissue may have infectious risks via prions.

As discussed, a good safety program requires both general universal precautions and precautions specific to certain pathogens or areas of concentration of a particular biorepository. In the United States, general training programs for blood-borne pathogens are required for all personnel who may potentially come into contact with human tissues. That training program should include education about how blood-borne and other pathogens are transmitted and how they cause human disease (see section on Training in Safety). In addition to blood-borne pathogens, the potential for exposure to non–blood-borne pathogens should be considered, such as tuberculosis, which is transmitted primarily via an airborne route by transfer of *Mycobacterium tuberculosis* that can be spread in tiny microscopic droplets released from infected tissue or fluids. Not all non–blood-borne pathogens are specifically regulated in every country. This is why the best safety programs not only follow regulations from local and nonlocal governing bodies but also consider their biorepository's unique circumstances and implement any additional safety measures intended to minimize risk of pathogen transmission to laboratory personnel; essentially, all dangerous organisms that are likely to be contacted by biorepository personnel should be addressed in any effective safety program.

DNA, RNA, and other molecular extractions from human and animal tissues may expose personnel to specific pathogens that can be transmitted to these employees. Commercial kits are frequently used to extract DNA and

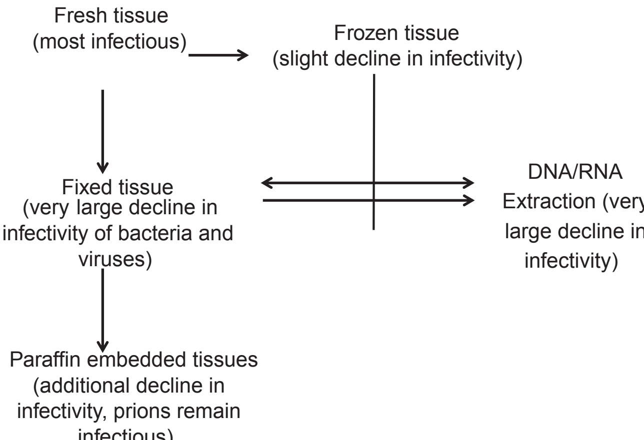


FIGURE 28.1 Effects of tissue processing on potential infection.

RNA from tissues. Extraction of nucleic acid including DNA from either human or animal tissue can contaminate the surfaces of a kit, equipment, or supplies with biohazardous agents that may be present in the human or animal tissues used for extraction. Therefore, universal precautions should be used during the extraction process and in handling cleaning and/or disposal of components of kits, equipment, and supplies.

As mentioned, one of the most vital components of any safety program is the appropriate use of PPE. The appropriate and intended use of PPE should be specified in the standard operating procedures (SOPs) of each task that a laboratory technician performs. The SOP should specify the proper use of such safety equipment including laboratory coats, gloves (e.g., appropriate size and type, which may vary depending on the specific task), and masks; protection from broken glassware and sharps such as needles and scalpel blades is important. All PPE, even laboratory coats, can be potentially misused and expose laboratory personnel to unnecessary risks. Of course, each task will have its own specific set of safety concerns that requires a specific set of PPE. This is one reason why SOPs are particularly useful in safety programs, because they should list the appropriate safety equipment for a particular task. A well-written SOP describes a task such that an employee can perform it simply by following instructions in the SOP. Potential exposure to small quantities of blood (e.g., 10-mL tubes) requires less PPE than does exposure to large volumes of blood such as during an autopsy. In the former example (handling a tube of blood), a laboratory coat, gloves, safety glasses, and mask may be sufficient to prevent exposure of skin and personal clothing to a pathogen, whereas in the latter example (during an autopsy), much more extensive PPE, such as surgical masks and scrubs, gloves, aprons, face shields, hairnets, and shoe covers, are appropriate to prevent exposure of pathogens to laboratory personnel. In some tasks, surgical-quality gloves should be used instead of lower-quality latex gloves because of potential exposure to bone shards and cutting equipment (scalpels or bone saws), which may increase the likelihood of infections via cuts or tears in gloves. Furthermore, specialized cutting gloves can be worn that are specially designed to prevent cuts from sharp objects. The use of respirators may be necessary in specific diseases. Of note, in the United States, use of respirators is governed by a specific OSHA regulation, (Occupational Safety and Health Administration 1910.134), including special training as well as fitting of respirators to each user.

In addition to the appropriate use of safety equipment including PPE, some biorepositories may elect to minimize exposure to potential pathogens by not even collecting or storing human or animal tissues that potentially contain or have been shown to contain blood-borne pathogens. Often this does not adversely affect specimen availability, because most investigators will not accept tissues that are likely to be contaminated with human pathogens. It has been the experience of the TCBF at UAB that not all personnel of laboratories to which biorepositories distribute tissues are adequately trained in biosafety. As such, it is recommended as part of the biorepository agreement or material transfer agreement (MTA) with the recipient institution that personnel of organizations or laboratories receiving tissue be appropriately trained in biosafety. Also, it may be appropriate to add an indemnification clause to the biorepository agreement or MTA to protect the biorepository legally in the event of injury to personnel at the recipient institution.

Finally, a biorepository's personnel must be aware that there are special national and international requirements for the transfer of human tissue and its derivatives, animal tissue, and infectious agents among organizations and across borders. The International Air Transport Association

(IATA) has specific training and shipping requirements. The regulations of the IATA and its training requirements cover biohazards as well as chemical hazards.

28.5.2 Chemical Hazards

Essentially all chemicals used in a biorepository may have adverse health effects, depending on the mode of contact and extent of exposure. Based on their effects, chemicals can be broadly characterized as toxins, irritants, allergens, carcinogens, and/or teratogens. In addition, some chemicals are potentially flammable and are considered fire hazards (e.g., xylene, alcohols, acetone) or even explosive risks, as discussed in the section on Fire Safety. Contact with chemicals can occur via several routes including direct contact with skin or mucous membranes (e.g., eyes, nose, mouth), inhalation via vapors or fumes, or even ingestion. Each mode of exposure has different types and extents of toxicity. One of the most common chemicals used to preserve tissue for research is the formaldehyde-based solution, formalin. If formalin were to come into contact with skin, it might induce irritation and/or an allergic reaction. If formalin vapors were inhaled, it could potentially have carcinogenic activity in addition to irritation and/or an allergic reaction to mucous membranes. Finally, if formalin were ingested, it could potentially be fatal.

Extraction of nucleic acids including DNA uses specific chemicals that must be considered hazardous. Because nucleic acid extraction is frequently performed using commercial kits, specific issues regarding safety as well as SDS are provided with the kits, usually by the specification of a website link. For chemicals used in the nucleic acid extraction process and not provided with the kit, SDS will be provided via the chemical supplier or chemical manufacturer. All other aspects of chemical precautions for extraction of DNA will be the same as for other laboratory uses of the chemicals.

As mentioned, an important component of a safety program should be devoted specifically to chemical safety, including proper handling and use of formaldehyde-based chemicals. The chemical safety plan should include training in chemical safety for all biorepository employees who could potentially come into contact with the chemicals. Also, items not typically associated with chemicals in the laboratory (e.g., paper or portable devices such as laptops, smartphones, or tablets) should be treated as being potentially contaminated.

The WHO, a special agency of the UN, has developed multiple components of a chemical safety plan and has newly implemented the GHS of the Classification and Labeling of Chemicals. The GHS requirement for labeling chemical containers consists of six major elements: product identifier, supplier identification, signal words, hazardous statements, precautionary statements, and pictograms.

Product identifier: product name and code (catalog number)

Supplier identification: manufacturer or supplier name, street address, city, state, postal code, and country

Signal Words: one of two words used to alert the user of the relative level of hazards severity of a product: "danger" and "warning." "Danger" is used for the most severe hazard categories. Only one single word can be used on a chemical label to indicate the severity of the hazard.

Hazard statements: phrases assigned to hazard classes and categories that describe the nature of a hazardous product, including, when appropriate, the degree of the hazard. The specific wording of a hazardous statement depends on the hazard classification. Examples of hazard statements include "Causes eye irritation," "Toxic if swallowed," and "Flammable liquid and vapor."

Precautionary statements: statements that describe recommended measures to be taken to minimize or prevent adverse effects resulting from exposure to a hazardous chemical. The statement may cover topics related to personal protection, storage, and disposal. Examples of precautionary statements include "Wear eye/face protection," "Store in well-ventilated place," and "If on skin, wash with plenty of water."

Pictograms: GHS has developed specific standard pictograms used in expressing chemical hazards. Pictograms are designed to indicate visually the chemical hazards associated with a particular chemical. Fig. 28.2 shows standard GHS pictograms used to label chemical containers and in SDS, with their associated descriptions and meanings.

An SDS serves as a workplace hazard warning, provides information to prevent direct exposure to chemicals, and specifies actions to be taken upon exposure to a specific chemical. As of June 1, 2015, the Hazard Communication Standard of the United States required chemical manufacturers, distributors, and importers to provide users with SDS to communicate in a uniform manner the hazards of each chemical provided. SDS formerly came in the form of material SDS, or MSDS. SDS are similar to MSDS except that SDS have a mandated, consistent, and user-friendly 16-section format. Sections 1–8 contain information on chemical and provider identification, hazards, composition, first aid measures, firefighting measures, accidental release requirements (spills and leaks), personal protection information, and proper handling, storage, and exposure controls. Sections 9–11 contain technical and scientific information, physical and chemical properties, stability and reactivity properties, and toxicologic information. Sections 12–15 contain ecological information, disposal considerations, transport information, and

Pictogram	Description	Chemical Hazards
	Flame over Circle	Oxidizer
	Flame	Flammables / Self-reactive chemicals; Emits flammable gas; Organic peroxides
	Exploding Bomb	Explosive; Self-reactive; Organic peroxides
	Skull and Crossbones	Acute toxicity (severe)
	Corrosion (metal or skin)	Corrosive
	Gas Cylinder	Gases under pressure
	Health Hazard	Carcinogen / Mutagenicity; Respiratory sensitizer; Target organ toxicity (acute and chronic); Aspiration toxicity
	Exclamation Mark	Acute toxicity; Severe responses; Target organ toxicity on a single exposure

FIGURE 28.2 Chemical hazard pictograms (container labels and safety data sheet).

regulatory information. Section 16, with "Other Information," is a general section that may include the date of SDS preparation, the last revision, and any other pertinent information. Thus, SDS provide valuable information for a biorepository regarding the hazards of each chemical, required personnel protection, emergency actions upon exposure, emergency contact information, and technical information that may affect the use, storage, transport, and disposal of a chemical.

Every chemical that is purchased commercially has an SDS that must be readily accessible by all laboratory personnel who use or could potentially come into contact with that chemical. The SDS is prepared by the manufacturer of the chemical or possibly a company that modifies the chemical (e.g., in case of repackaging). Biorepositories outside the United States may follow different regulations, although most SDS are now international, and associated chemical hazards are often regulated similarly.

Although each chemical has a specific SDS with specific hazards and safety regulations, combinations of chemicals used in a laboratory may create new hazards that are different from those indicated in each individual chemical's specific SDS. As such, combining chemicals in a laboratory may require special precautions and safety procedures, and may be prohibited in some instances. For example, it is well-known that combining concentrated strong acids (e.g., HCl) with strong bases (e.g., NaOH) without appropriate dilution can create a hazardous reaction, but it is less known that combining strong oxidizers (e.g., potassium permanganate) with high carbon–containing materials (e.g., ethylene glycol) can also cause a hazardous situation. The practice of combining chemicals should be strictly regulated in a laboratory and approached with great care. This should adequately be addressed in the educational component of the safety program.

As with biohazards, employees should communicate with their supervisors regarding their knowledge of the safety program and how it applies to their experiences in the workspace in which they handle chemicals. The GHS has already had a large impact on training in chemical hazards and influences the modification of training as new mandated formats for the labeling of chemicals and use of SDS are implemented. In addition, as with biohazards, an appropriate engineering infrastructure and proper use of PPE are necessary to prevent injury owing to chemicals. Nearby eyewashes, a shower, and easily accessible fire extinguishers may be particularly relevant in parts of the laboratory where some chemicals are used. Latex gloves are not appropriate protection against chemical exposure.

One component of proper engineering infrastructure pertains to the adequate storage of chemicals, especially flammable chemicals and caustic agents, which should be kept in certified chemical storage cabinets. Ultimately, storage of large quantities of flammable chemicals should be minimized or avoided; furthermore, certain types of chemicals should be stored separately, such as strong acids and strong bases. In addition, a storage cabinet that contains flammable chemicals, if practicable, should not be positioned near an entrance or exit of a laboratory space. In some cases, for particularly toxic chemicals or chemicals with associated dangerous vapors, even more extensive engineering practices should be implemented. All of these considerations are part of the design and appropriate application of an engineering infrastructure, and make up an important component of a biorepository's safety program.

Specific requirements and regulations related to chemical safety in the United States and some other countries can be found in Occupational Exposure to Hazardous Chemicals in Laboratories (29 CFR, 1910.1450), the Hazard Communication Standard (29 CFR, 1910.1200), and the Formaldehyde Standard (29 CFR, 1910.1048). Similar regulations for laboratories in the EU include OSHA Europa 94/24/EE, which sets minimum requirements for the protection of workers exposed to chemical agents in the workplace. 2004/37/EEC protects workers from the risks related to exposure to carcinogens and mutagens at work. 91/322/EEC, 96/94/EC, 2000/39/EC, and 2006/15/EC establish occupational exposure limits for the EU. 92/58/EEC contains the regulations on classification and labeling of chemicals and provides important information concerning safety labels of chemical containers and SDS.

In the United States, the Occupational Exposure to Hazardous Chemicals in Laboratories (the Laboratory Standard) requires that employers develop a written chemical safety plan. The chemical safety plan for a biorepository must be able to protect all employees from hazardous chemicals. Specifically, the chemical safety plan should maintain chemical exposures below the action level or, in its absence, the permissible exposure limit for each specific chemical. Biorepository safety plans also should clarify and incorporate, if appropriate, the Formaldehyde Standard, especially if the biorepository repackages or prepares formaldehyde solutions and provides them outside the biorepository. The minimum required elements of a chemical safety plan should include the following:

1. plans for developing and implementing the chemical safety plan and monitoring its use and effectiveness
2. plans for monitoring the exposure of employees or visitors to hazardous chemicals, reporting and documenting chemical safety incidents, and modifying the safety plan to prevent recurrence
3. an effective training program for all individuals working with chemicals or in areas where chemicals are used or stored
4. a written emergency plan to address chemical spills. The plan should incorporate prevention, containment, cleanup, and waste disposal, including contaminated materials used in a cleanup and approaches to minimize personnel exposed during cleanup
5. a policy to minimize exposure to hazardous fumes, which includes minimizing inhalation by personnel, monitoring the effectiveness of ventilation of any space where volatile chemicals are used or stored, and planning for ventilation failure
6. a policy that includes minimizing exposure of skin, eyes, mucous membranes, and personal clothing to chemicals and washing of hands and exposed skin after use and/or exposure to hazardous chemicals
7. plans for emergency evacuation and medical care for employees injured by hazardous chemicals and safety drills for chemical emergencies

8. policies prohibiting eating, drinking, smoking, gum chewing, application of cosmetics in the laboratory, and storing food and/or beverages in laboratory storage areas, laboratory refrigerators, and freezers
9. a policy requiring protection of personnel and visitors from chemical exposures. This includes wearing appropriate protective equipment including laboratory coats, suitable gloves, and eye and mucous membrane protection when there is a potential for contact with toxic chemicals. Gloves should be appropriate for chemical protection and should be inspected before use and washed before removing. Mouth pipetting and mouth suctioning for starting a siphon must be prohibited. The use of contact lenses in the laboratory should be avoided
10. a policy to document and observe chemical exposure limits, if established, and to follow legal requirements for specific hazardous chemicals. Employees should be knowledgeable regarding these issues

Chemical safety should be a priority for biorepositories, whether they use large or small quantities of chemicals. For example, biorepositories that are largely concerned only with production and staining of microscopic slides of tissues may use picric acid in some of their histochemical fixatives and staining procedures; of importance, if picric acid is permitted to dry, the chemical constitutes an explosive hazard. As noted, combining even small amounts of chemicals can be extraordinarily dangerous and possibly result in spontaneous combustion or accelerated heating, which may cause injuries to laboratory personnel as a result of liquids splashing or boiling over. In general, appropriate approaches for the use of chemicals in the laboratory should include the following:

1. All solutions made in the laboratory must be labeled, including ingredients and hazards as well as the day prepared, expiration date, and preparer (initials are adequate). Combinations of chemicals should include the SDS for each component chemical. In addition, laboratory personnel must assume that a combination of chemicals is potentially more hazardous than simply the sum of the toxicities of the individual chemical components.
2. If a chemical substance is produced or repackaged by a laboratory for outside use, the laboratory is required to prepare an SDS for the other site(s), which includes the new GHS labeling requirements.
3. Regarding handling of chemicals, prudent approaches include the following:
   - **a.** Follow the chemical safety plan.
   - **b.** Contact with chemicals should be avoided, and chemicals should be washed quickly from exposed sites.
   - **c.** Appropriate safety equipment (e.g., laboratory coat, safety glasses, mask, gloves) should be used when handling chemicals.
   - **d.** Chemicals with no known significant hazards should also be handled with care.
   - **e.** Chemicals with moderate to severe hazards should be handled with special precautions.
   - **f.** Care should be taken during the maintenance of equipment or other surfaces that could potentially be contaminated with hazardous chemicals.

Many approaches to chemical and the other areas of safety are regulated in the United States by OSHA of the Department of Labor. In the United States, OSHA can inspect biorepositories and impose large fines for safety violations.

Chemicals common to biorepositories include formaldehyde, xylene, acetone, and alcohols. Formaldehyde is the most common type of fixative used in the preparation of paraffin blocks. Formaldehyde is an irritant, allergen, and carcinogen. Outside biorepositories, formaldehyde is used in many products including furniture and carpets. An allergy to formaldehyde can be debilitating. In the United States, there is a special law that addresses the use of chemicals containing formaldehyde (Formaldehyde Standard 29 CFR, 1910.1450). Xylene, acetone, and alcohols are hazardous because of their flammable vapors and toxicity.

28.5.3 Physical Security and Safety

Essentially all laboratories, including biorepositories, have physical safety considerations. Physical safety incorporates a wide range of concerns, from ensuring employees are not physically injured by others to preventing falls. Ensuring physical safety begins with the quality of the overall infrastructure (e.g., facilities). Damaged steps or rugs, slippery floors, and inappropriate use of furniture and equipment can lead to all sorts of injuries to employees. Physical safety considerations may also be specific to certain functions of the laboratory. New standards for areas of health care facilities that store liquid nitrogen and other cryogenic fluids have been developed by the NFPA. These include NFPA.55 and NFPA.99. Standards such as these affect the design and/or infrastructure of biorepositories, and individuals belonging to safety committees that oversee laboratories should stay up-to-date on these types of developments.

Some common examples of poor compliance with physical safety requirements include using chairs as ladders, slippery floors contaminated by paraffin and soaps or fluids, and file cabinets, cylinders, and shelves that are inadequately secured; all of these issues can potentially lead to injuries. Unsecured equipment such as file cabinets and other furniture are especially dangerous during earthquakes. Employees may also experience physical injuries

such as back trauma as a result of lifting (e.g., racks from liquid nitrogen freezers), repetitive action–type injuries, or burns from both hot or cold (e.g., from dry ice or liquid nitrogen) substances. Minor injuries such as minor cuts (e.g., from paper) are more difficult to avoid. Minor injuries should not be exacerbated by exposure of broken skin to toxic chemicals or biohazards. All of these issues should be addressed in a biorepository's safety program.

General security of a biorepository requires limiting or preventing access by unauthorized personnel. Also, threats to employees, especially by other employees, should not be tolerated. Consideration of employees' safety may extend outside the biorepository walls and include emergency exits and pathways and areas surrounding the workplace, including parking areas.

28.5.4 Fire Safety

Biorepositories may store and use chemicals that may be fire hazards, such as xylene, acetone, and ethanol. Thus, fire safety should be of concern to such biorepositories. One goal of the fire safety program should be to facilitate personnel rapidly escaping from fires and other emergencies; thus, clear emergency exit pathways should be posted at the exits of all rooms as well as general areas of the biorepository. The emergency exits should be described in the safety training program. Emergency exits should never be blocked, obstructed, or locked; hallways should not be cluttered or obstructed. Similarly, access to fire blankets, showers, and fire equipment must not be impeded. Fire equipment (e.g., showers and fire extinguishers) should be tested periodically and the results documented. There should be periodic fire drills to ensure that emergency exits are known and accessible. Flammable chemicals should be stored appropriately in chemical safety cabinets. The volume of flammable agents stored in each cabinet should not exceed the cabinet's protective standard, and the total amount stored in biorepositories should be appropriate and may be regulated locally and/or by accreditation and certification agencies.

Smoking should be prohibited in the biorepository. Furniture, rugs, and equipment should meet fire retardant standards. Similarly, doors that serve as barriers to fire should meet the requirements for construction of buildings that house laboratories such as biorepositories. In the United States, aids to minimize fire hazards are available from the NFPA. One useful approach to evaluating fire safety issues is to invite an inspection by the local fire department.

28.5.5 Electrical Safety

All equipment of the biorepository must be electrically grounded, and the grounding should be tested for each piece of equipment when it is put into operation, and at least yearly thereafter. Electrical base plugs must also be grounded (and testing documented) and also should be in good working condition. All work on the electrical system should be approached with great care by certified electricians who should ensure that all electrical working areas are protected by removal or inactivation of fuses; there should be written warnings at the fuse box to prevent activation while electrical work is in progress.

Each site should decide whether personal electrical devices can be brought into the biorepository. If permitted, electrical appliances such as radios should not be inadvertently ignored when testing biorepository equipment for grounding. In addition, personnel should be careful with electrical appliances and equipment around water, especially bathrooms or showers and sinks.

28.5.6 Other Areas of Safety

Other safety concerns include specialized areas of safety such as radiological safety. Discussions of such safety areas are included in specialized institutional safety programs and are beyond the scope of this chapter.

28.6 CONCLUSIONS

The safe operation of a biorepository requires all employees to work to maintain a safe working environment. All employees of the biorepository need to contribute to developing and monitoring an effective safety program. The infrastructure of the biorepository should be monitored by the SO/safety committee, and problems should be identified and corrected rapidly; training of all personnel as to safety hazards should be appropriate for special types of biorepository operations including protecting against biohazards, chemical hazards, and physical hazards; training should generally be updated at least yearly and this should be documented. Also, biorepositories should identify and monitor electrical and fire hazards. The safety plan should also use engineering practices to aid in ensuring a safe working environment. This includes the proper use of safety equipment such as laboratory coats, safety glasses, masks, and appropriate gloves, as well as good ventilation and drainage of the site. The safety plan should be dynamic and carefully evaluated, with annual reviews and changes necessary to minimize the recurrence of all safety incidents, which should be evaluated carefully.

ACKNOWLEDGMENT

This work has been supported by the SPORES at UAB in Pancreatic Cancer (5P20CA101955-05), and Breast Cancer (2P50CA089019-06A2); the Tissue Procurement Shared Facility of the Comprehensive

Cancer Center (P30CA-13148-40); the Cooperative/Collaborative Human Tissue Network, Southern Division (5U01CA44968-18S2); and the Pulmonary Hypertension Breakthrough Initiative of the National Heart, Lung, and Blood Institute (1R24HL123767-01) of the National Institutes of Health and the Cardiovascular Medical Research and Education Fund of the American Heart Association.

REFERENCES

Aamodt, R.L., Anouna, A., Baird, P., Beck, J.C., Bledsoe, M.A., DeSouza, Y., Grizzle, W.E., Gosh, J., Holland, N.T., Hakimian, R., Michels, C., Pitt, K.E., Sexton, K.C., Shea, K., Stark, A., Vaught, J., 2005. Best practices for repositories I: collection, storage and retrieval of human biological materials for research. Cell Preserv. Tech. 3, 5−48 first ed.

Astrin, J., Baker, S., Barr, T.J., Benson, E., Cada, M., Campbell, L., Campos, A.H.J.F.M., Carpentieri, D., Clement, O., Coppola, D., De Souza, Y., Fearn, P., Feil, K., Garcia, D., Giri, J., Grizzle, W.E., Groover, K., Harding, K., Kaercher, E., Kessler, J., Loud, S., Maynor, H., McCluskey, K., Meagher, K., Michels, C., Miranda, L., Muller-Cohn, J., Muller, R., O'Sullivan, J., Pitt, K., Pugh, R., Ravid, R., Sexton, K., Silva, R.L.A., Simione, F., Skubitz, A., Somiarai, S., van der Horst, F., Welch, G., Zaayenga, A., 2012. Best practices for biorepositories: collection, storage, retrieval and distribution of biological materials for research. Biopreserv. Biobank. 10, 79−161. http://dx.doi.org/10.1089/bio.2012.1022 third ed.

Beekmann, S.E., Doebbeling, B.N., 1997. Frontiers of occupational health. New vaccines, new prophylactic regimens, and management of the HIV-infected work. Infect. Dis. Clin. North Am. 11, 313−329.

Cardo, D.M., Bell, D.M., 1997. Bloodborne pathogens transmission in health care workers. Risks and prevention strategies. Infect. Dis. Clin. North Am. 11, 331−346.

Grizzle, W.E., Fredenburgh, J., 2001. Avoiding biohazards in medical, veterinary and research laboratories. Biotech. Histochem. 76, 183−206.

Grizzle, W.E., Polt, S.S., 1988. Guidelines to avoid personnel contamination by infective agents in research laboratories that use human tissues. J. Tissue Cult. Methods 8, 191−200.

Grizzle, W.E., Bell, W.C., Fredenburgh, J., 2010. General considerations concerning safety in biomedical research laboratories. In: Patrinos, G.P., Ansorg, W. (Eds.), Molecular Diagnostics, second ed. Elsevier B.V., New York, pp. 563−572.

Grizzle, W.E., Gunter, E.W., Sexton, K.C., Bell, W.C., 2015. Quality management of biorepositories. Biopreserv. Biobank. 13 (3), 183−194. PMC26035008.

NIH Guidelines for Research Involving Recombinant or Synthetic Nucleic Acid Molecules, November 2013. http://osp.od.nih.gov/office-biotechnology-activities/biosafety/nih-guidelines (accessed 25.04.16).

Occupational Safety and Health Administration, Personal Protective Equipment 1910.134, Subpart 1: Respiratory Protection, United States Department of Labor. Available via: https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=STANDARDS&p_id=12716 (accessed 02.12.15).

Occupational Safety and Health Administration, Hazard Communication Standard 1910, Subpart Z: Safety Data Sheets. United States Department of Labor. Available via: https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=STANDARDS&p_id=10103 (accessed 02.12.15).

Occupational Safety and Health Administration, OSHA Aligns Hazard Communication Standard with the United Nations' Globally Harmonized System of Classification and Labeling of Chemicals, United States Department of Labor. Available via: https://www.osha.gov/dsg/hazcom/ (accessed 02.12.15).

Occupational Safety and Health Administration, Hazard Communication Standard: Labels and Pictograms, United States Department of Labor. Available via: https://www.osha.gov/dsg/hazcom/index.html (accessed 02.12.15).

Pitt, K., Campbell, L., Skubitz, A., Somiari, S., Sexton, K., Pugh, R., Aamodt, R., Baird, P., Betsou, F., Cohen, L., De Souza, Y., Gaffney, E., Geary, P., Grizzle, W.E., Gunter, E., Horsefall, D., Kessler, J., Michels, C., Kaercher, E., Morales, O., Morente, M., Morrin, H., Petersen, G., Robb, J., Seberg, O., Thomas, J., Thorne, H., Walters, C., Riegman, P., 2008. Best practices for repositories: collection, storage, retrieval and distribution of biological materials for research. Cell Preserv. Tech. 6, 5−58 second ed.

Chapter 29

Quality Assurance in Genetic Laboratories

E. Dequeker
University of Leuven, Leuven, Belgium

29.1 INTRODUCTION

Medical laboratories, and specifically genetic testing laboratories, provide vital medical services to different clients: clinicians requesting a test, patients from whom the sample was collected, public health and medicolegal instances, referral laboratories, and authoritative bodies. The field of medical genetics has witnessed a remarkable rate of progress both in the understanding of the association between genetic variants and human disease, and in the technical ability to perform genetic analysis. As a consequence, there is a huge growth in demand from patients and medical health professionals for clinical genetic testing (Dequeker, 2010).

All expect results that are accurate and obtained in an efficient and effective manner, within a suitable time frame and at acceptable cost. There are different ways to achieve the end result. Accreditation of health care services is recommended to ensure operation to the highest quality standards. Various initiatives, such as legislation or accreditation by the International Organization for Standardization (ISO) or Joint Commission International may be active to support and improve quality. Compliance with ISO 15189, the international standard for the accreditation of medical laboratories, is becoming progressively accepted as the optimal approach to ensuring quality in medical testing (Berwouts et al., 2010; Tembuyser et al., 2016).

In the early 2000s (Berwouts et al., 2012a,b), concerns were raised with regard to quality assurance practices in Europe and beyond because of the rapid expansion of genetic testing (McGovern et al., 1999; Dequeker et al., 2001; McGovern et al., 2003). The European Science and Technology Observatory Network and the Organization for Economic Cooperation and Development (OECD) conducted studies to determine the state of quality systems within genetic testing services, identifying an urgent need for better and more consistent quality assurance practices (European Science and Technology Observatory Network, 2003; Organisation for Economic Cooperation and Development, 2005). As a result, initiatives were taken to improve the quality of genetic testing laboratories. External quality assessment (EQA) providers such as the Cystic Fibrosis Network and the European Molecular Genetics Quality Network received support from the European Commission (BMH4-CT96-0462, QLK3-CT99-00,241, and SMT4-CT98-7515) (Dequeker and Cassiman, 1998; Muëller, 2001). Another initiative was the CRMGEN project, which aimed to develop certified reference materials for molecular genetic tests (G6RD-CT-2001-00,581) (Barton et al., 2004). In 2005, a network of networks was established, named EuroGentest (Cassiman, 2005). After 2 years, the OECD published Guidelines for Quality Assurance in Molecular Genetic Testing (OECD, 2007). In the United States, the Centers for Disease Control and Prevention (CDC) formulated recommendations for improving quality assurance of molecular genetic testing laboratories, and efforts were made to create a genetic testing specialty under the Clinical Laboratory Improvement Amendments (Williams et al., 2003; United States Department of Health and Human Services 2011).

There is an understandable assumption from users that the results of genetic tests are reliable, whereas studies of performance in EQA schemes repeatedly show that errors occur at a measurable rate at all stages of the testing process (Seneca et al., 2008; Touitou et al., 2009). To respond to the need to improve the quality of clinical

Molecular Diagnostics. http://dx.doi.org/10.1016/B978-0-12-802971-8.00029-8
Copyright © 2017 Elsevier Ltd. All rights reserved.

genetic testing, providers are encouraged to implement a comprehensive quality management system (QMS) and to fulfill the requirements for internationally recognized standards for laboratory accreditation (Ibarreta et al., 2004).

29.2 INTERNATIONAL STANDARDS

During the typical laboratory accreditation process, the laboratory is assessed against a "standard document" that contains a number of specifications equivalent to the clauses in a normative document. Other accreditation systems used documents that describe guidelines, principles, or criteria as the basis for assessment (Dequeker, 2010; Burnett, 2002).

The world's largest developer of standards is the ISO, a nongovernmental network of the national standards institutes of 162 countries. In Europe, another nongovernmental organization, the European Committee for Standardization (CEN), is composed of 39 national standardization institutes. Since the Vienna Agreement was signed in 1991, these two international standard organizations have worked together closely by mutual agreement with the national standardization institutes. This agreement endorses, among many other engagements, an exchange of technical information between CEN and ISO, the adoption in Europe of existing ISO standards without text changes, and the parallel approval of newly developed documents within both organizations. All of the approved CEN standard should be adopted in each member state of the European Union within 6 months after approval, and should replace all (related) national standards (Dequeker, 2010).

Each standard is assigned a unique identification number, often followed by the year of approval or revision. In front of the unique number, the acronym of (sometimes multiple) standardization institutes can be found that adopted the standard.

Major standards relevant to quality management in clinical genetic laboratories are summarized in Table 29.1 (Dequeker E, 2010). The ISO 9001, first published in 1987, is the oldest and most generic standard. ISO defines "quality management" as the comprehensive set of measures that is taken by an organization to (1) meet customers' quality needs, (2) meet applicable regulatory requirements, and (3) continually improve performance. The quality system "requirements" in ISO 9001 are universal, i.e., they can be applied to any organization (private business enterprises, public nonprofit organizations, or government departments) independent of the organization size and irrespective of the nature of the product manufactured, service delivered, or analytical test performed. Briefly summarized, ISO 9001 concerns the way activities in an organization are performed, but does not assess the actual results of these activities (at least, not directly). Whereas ISO 9001 describes the overall requirements for any quality system, it does not dictate details regarding specific criteria that should be fulfilled, which leaves plenty of room for flexibility when it is implemented in particular organizations from different business sectors and varying national cultures. This is a typical characteristic for all ISO standards.

The ISO has developed an international standard for the accreditation of testing laboratories in general (ISO 17025) and one specifically for medical laboratories (ISO 15189) (Dequeker, 2010; Canadian Standards Association, 2004; International Organization for Standardization: ISO 15189 medical laboratories, 2012). ISO 15189 emphasizes the quality of contributions to patient care, as well as that of laboratory and management procedures, and is therefore the preferred standard for genetic testing laboratories (Kenny, 2001; Burnett, 2006; Burnett and Blair, 2001; Harper et al., 2010). In contrast, ISO 17025 is written in more general terms and is applicable to a wide range of testing environments. However, because the ISO 15189 standard is designed for a wide range of medical laboratories, a certain effort of reflection and interpretation may be necessary to apply it in the context of genetic testing. This standard covers all different aspects of medical laboratory activities, including the provision of advice to customers of the laboratory service, the collection of patient samples, the interpretation of test results, acceptable turnaround times, and the laboratory's role in the education and training of health care staff. In addition, this standard specifies requirements for the preanalytical phase as well as for a demonstration of the clinical relevance of a particular test. Currently, most accredited medical laboratories in Europe have adopted this standard or are in the process of doing so (Dequeker, 2010).

Apart from accreditation standards, the OECD has published specific Guidelines for Quality Assurance in Molecular Genetic Testing that are equally applicable to cytogenetics and biochemical genetics. This OECD guideline document might be considered a sector-specific document to be used in combination with existing accreditation standards (Berwouts et al., 2010; OECD, 2007). The contents of the guidelines are not formal requirements but they can be a useful complement to laboratories for improvement and harmonization. The minimum common requirements described address general principles and best practices, quality assurance, external quality assessment, reporting of results, and training for laboratory personnel.

Personal copies of ISO standards can easily be bought through the website of ISO (http://www.iso.org) or through the National Standards Organization [e.g., Bureau voor Normalisatie (NBN) in Belgium, Deutsches Institut für Normung (DIN) in Germany, and Association Française de Normalisation

(AFNOR) in France]. An overview of these bodies can be found on the website of the European Committee for Standardization, at http://www.cen.eu/cenorm/members/national+members/ (Berwouts et al., 2010).

In the United States, the Clinical Laboratory Standard Institute published policies, processes, and procedures regarding laboratory QMS, including detailed guidelines and recommendations for technologies in clinical genetic testing services (CLSI, 2012). Of note, accreditation conforming to ISO 15189 is rare in US laboratories. Rather, US laboratories follow the local requirements of the Clinical Laboratories Improvement Act and the checklist of the College of American Pathologist (CAP) for certification. Finally, the OECD (OECD, 2007) published guidelines for Quality Assurance in Molecular Genetic Testing that are equally applicable to cytogenetics and biochemical genetics.

29.3 ACCREDITATION AND CERTIFICATION

There is still a misunderstanding about the difference between accreditation and certification (Berwouts et al., 2010). Certification is a procedure by which a third party gives written assurance that a product, process, or service conforms to specific requirements. These specifications are written by the laboratory or the manufacturer itself that applies for the certificate. Furthermore, any certified third party (private company, government, etc.) that is fully independent from the certification requester can make the assessments necessary for the certification procedure. Consequently, a certification does not guarantee a comparable quality between two different businesses or services. The ISO 9001 standard is the best known certification standard. Certification of an organization's QMS against this standard will confirm the compliance to this standard, but it does not assess the specific technical competence of a laboratory. This means that ISO 9001 certification of a laboratory does not assure accurate, reliable, or state-of-the-art analytical testing (Dequeker, 2010).

Accreditation is the procedure by which an authoritative body formally recognizes that a laboratory is competent to carry out specific tasks. This goes way beyond certification, because (1) an accreditation certificate cannot be delivered by any third party, and (2) it is a quality system that is composed of certain technical and analytical requirements. Typical standards for laboratory accreditation are ISO 17025 and ISO 15189 (Table 29.1). Accreditation obliges laboratories to meet a predefined set of standards regarding laboratory management, facilities, staff qualifications and training, performance, and regular participation in external quality assessments. Accreditation of a laboratory can be specific (e.g., certifying competence in carrying out a single specific test) or broad (e.g., covering all analytical testing performed in a laboratory) (Dequeker, 2010).

For many years, accreditation was predominantly seen as a voluntary activity. However, accreditation for clinical genetic testing laboratories has been widely embraced by governments and policy makers in a growing number of countries (OECD, 2007). Accreditation for genetic testing laboratories is mandatory in a number of countries, such as Belgium and France. The demand for accreditation certificates given to laboratories by a recognized (international) accreditation body will help to ensure compliance with upfront set requirements from government and regulators, and may also serve as a safeguard for the public.

Of note, some countries implemented a so-called licensing system for health care facilities. This is distinct from accreditation and certification, and is usually mandatory and government-imposed (e.g., "agrément" in France). Licensing does not necessarily require the evaluation of quality management or technical competence (Dequeker, 2010).

29.4 ELEMENTS OF A QUALITY MANAGEMENT SYSTEM

The initial step of introducing good-quality management in the laboratory is the identification of key elements of a quality system. These elements need to be integrated with the existing path of workflow, processes, and organization through the documentation of standard operating procedures (SOPs) and the definition of objectives and policies in a quality manual (Berwouts et al., 2010). Laboratory activities to address these needs include planning and preparation activities to ensure the laboratory is ready to introduce or provide genetic testing services, and activities to validate or verify the performance of new or updated examination procedures. These activities are prerequisites for providing examination services to laboratory users (CLSI, 2012) (Fig. 29.1).

To support laboratories working toward accreditation and improve the understanding of quality assurance, interactive workshops were developed within the framework of the EuroGenTest Network of Excellence (EUGT NoE, FP6-512,148) (Berwouts et al., 2010; Cassiman, 2005). A successful format was achieved by integrating expertise in both laboratory and quality management, and in learning and change management. Further collaboration with the educational committee of the European Society of Human Genetics will guarantee the continuity of these workshops.

All essential elements of a QMS are covered by the ISO 15189/ISO 17025 accreditation standards (Fig. 29.2). This international standard specifies the requirements for both

TABLE 29.1 Standards Relevant to Quality Management in Clinical Genetics Laboratories

Standard	Title	Example Applications in Medical Genetics
ISO 9001	Quality management systems requirements	Certification of compliance for genetic counseling services or clinical consultations.
ISO/IEC 17025	General requirements for competence of testing and calibration laboratories	Accreditation of testing laboratories.
		ISO 15189 is generally more appropriate for medical testing laboratories.
ISO 15189	Medical laboratories requirements for quality and competence	Accreditation of testing laboratories. Pre- and postanalytical procedures, including patient contact, information, and sampling can be included.

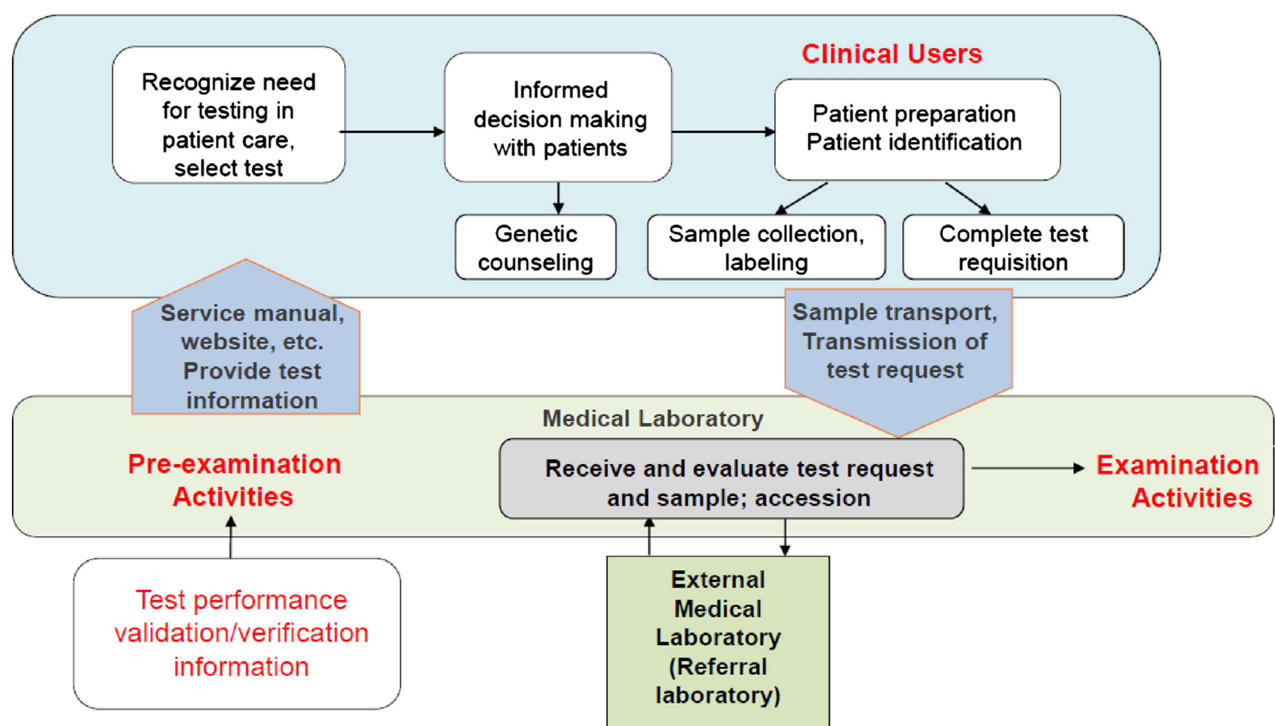


FIGURE 29.1 The laboratory's path of workflow for genetic testing and its interfaces with clinicians and other users of laboratory services (CLSI, 2012).

quality and technical/scientific competence in medical laboratories. Management aspects include document control; identification of nonconformities; implementation of corrective and preventive actions; and action plans to continuously improve performance of internal audit and management review, and resolution of complaints and evaluation of external services, suppliers, contracts, and referral laboratories. Technical elements enclose personnel and training, accommodation, equipment, and validation, and assuring quality of examination procedures by internal quality control (IQC), EQA, maintenance, and calibration (Berwouts et al., 2010). The most important parameters of a quality system and their practical implementation in the laboratory are discussed subsequently.

29.4.1 Organization and Quality Management

The genetic testing laboratory, or the organization of which the laboratory is a part, must be legally identifiable. The laboratory must define "who is legally responsible for the work of the laboratory in the event of legal action being taken." Confidentiality is also critical in genetic testing services, and policies on this matter should be written down (Berwouts et al., 2010).

All responsibilities and authorities of personnel should be defined, and as a result, conflicts of interest should be identified. These could include financial, commercial, and research or other influences, external or internal. The

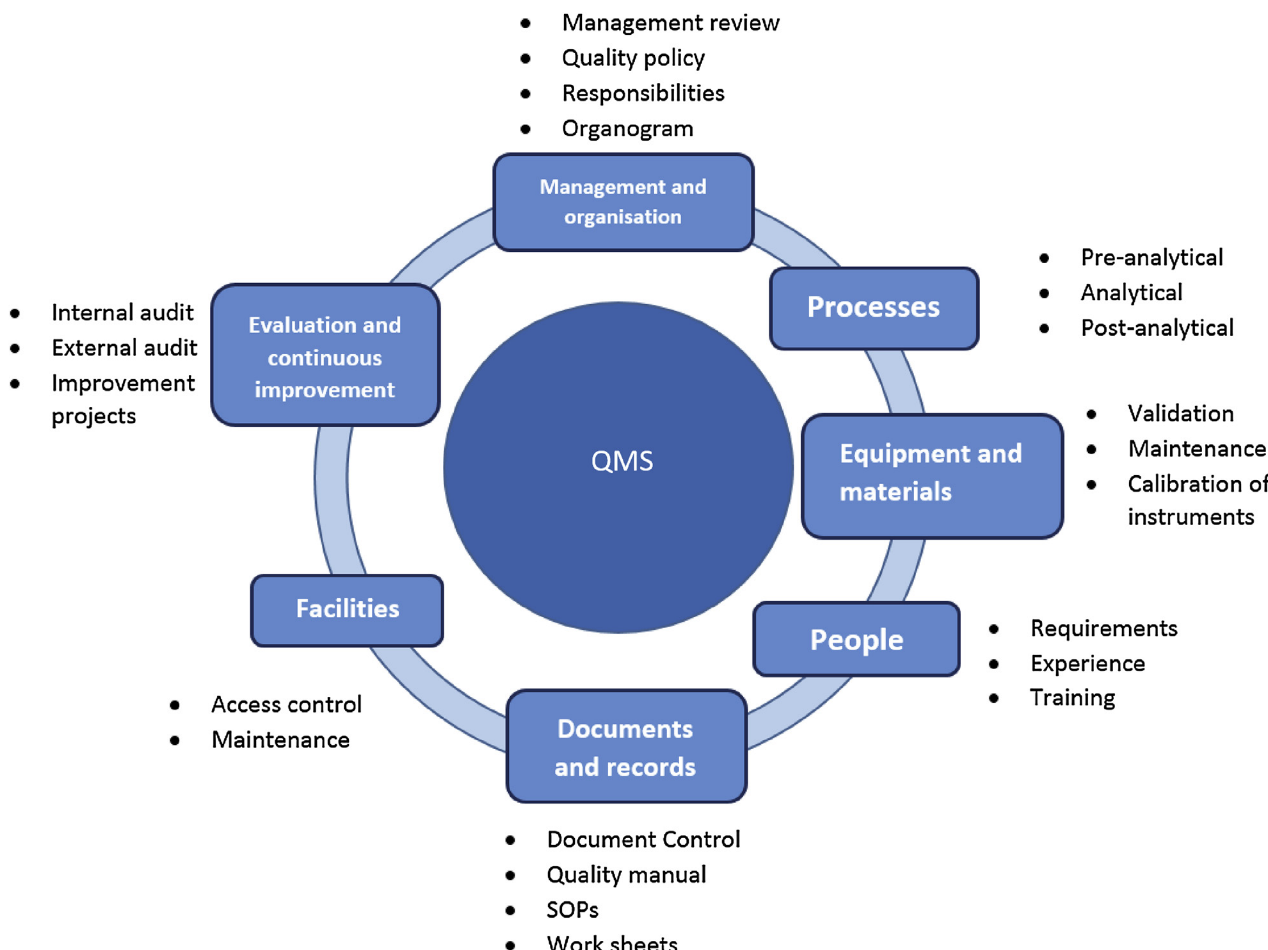


FIGURE 29.2 Elements of a quality management system (QMS). *SOP*, standard operating procedure.

design, implementation, maintenance, and improvement of the QMS are the responsibilities of laboratory management. Therefore, they must be provided with sufficient time and resources to carry out their duties, which may lead to the reorganization of certain positions. The quality manager is the person who takes overall responsibility for the QMS and reports directly to the level of laboratory management. However, it is important for the quality manager not to work in isolation, but for him or her to interact with and get support from other people working in the laboratory with respect to the QMS. Typically, laboratories set up a quality team to divide work and get everybody involved. Finally, personnel should receive adequate training and supervision, depending on their experience and level of responsibility. The whole QMS must be described in a quality manual. This quality manual is the formal, top-level document that establishes the quality policies and objectives of an organization. It is used internally to outline the format of documents used in the QMS, as well as the roles and responsibilities of the personnel responsible for management of the QMS or technical procedures, and also externally to give a clear and precise view of the organization and activities of the laboratory to potential clients and inspectors. ISO 15189 provides clear instructions on the content of the quality manual and quality policy (Berwouts et al., 2010).

29.4.2 Document Control

Several documents will be generated for a QMS: SOPs, worksheets, log books, validation and training files, and a quality manual. These documents must be created, implemented, communicated, and understood by all relevant personnel. Document control is the mechanism by which QMS documents are created, approved, distributed, reviewed, revised, and archived. In other words, all documents are approved by authorized personnel before use and a distribution list is available that identifies the current valid versions.

Documents are periodically reviewed or revised when necessary, and then reapproved. Invalid or obsolete documents are removed and the archived documents are

recognizable to ensure that staff uses only the latest authorized versions. Finally, procedures must be defined according to whether and how amendments and minor and major changes, including handwritten changes, can be made. Fig. 29.3 is a schematic overview of the different phases in document control (Burnett, 2002).

29.5 QUALITY CONTROL

IQC is an internal verification that the test yields consistent results day after day; in other words, it is the identification measure of precision, but not necessarily of accuracy. ISO 15189 requires that "the laboratory shall design IQC systems that verify the attainment of the intended quality of results," but it does not provide details of specific controls (Berwouts et al., 2010; Dequeker, 2010). On the one hand, the laboratory should avoid mistakes (ISO 15189, 5.6.1) in the process of handling samples, requests, examinations, reports, and so on; on the other, the laboratory should determine uncertainty (ISO 15189, 5.6.2) when relevant and possible. For each test, the laboratory should identify and define the potential errors, risks, and challenges (typically, during the validation phase); subsequently, specific IQC should be defined to ensure each risk and potential problem. The defined controls should be documented in SOPs (Westgard, 2003, 2006). Errors can also be traced and reduced by keeping records of batch numbers of all laboratory solutions to improve traceability and troubleshooting. Labeling each tube with a unique code and double checking before, during, and after transferring samples is also a preventive IQC measure. The standard states, "There shall be effective separation between adjacent laboratory sections in which there are incompatible activities."

29.6 QUALITY ASSESSMENT

Although genetics laboratories are often in close contact with each other, they operate in isolation and rarely compare or benchmark their data with others. Therefore, EQA or proficiency testing provides an opportunity to undertake such comparisons and to have an independent appraisal of the laboratory's data compared with reference values or performance criteria, or with the performance of similar laboratories (Dequeker, 2010). Results from participation with EQA provides confidence to the laboratory director that the laboratory's performance is satisfactory or otherwise alerts that investigation of potential problems within laboratory is required. Several organizations initiated EQA schemes of different scales and for different genetic disorders. Schemes are organized by international groups such as the European Molecular Quality Network, Cystic Fibrosis Network, Cytogenetics External Quality Assessment scheme, European Research Network for evaluation and improvement of screening, Diagnosis and treatment of Inherited Disorders of Metabolism, and by national groups such as the UK National External Quality assessment. In the United States, CAP and CDC are the most recognized. Currently, laboratories from other continents most often join the European or US

PREPARE
A new document should be created by competent personnel. The draft document is uniquely identified.

APPROVE
The document moves from a draft to being an active document and is authorized by designated staff before being issued.

ISSUE AND DISTRIBUTE
Date of issue and version number should be added. Distribution can electronically or on paper. A distribution list should be maintained.

REVIEW
Reviewing a document is to ascertain its continuing 'fitness for purpose'. Changes should be dated and authorized.

REVISE AND ARCHIVE
In case of major changes, the document should be revised and a new version issued. Obsolete documents are removed and archived.

FIGURE 29.3 Document control process. *From Berwouts, S., Morris, M.A., Dequeker, E., 2010. Approaches to quality management and accreditation in a genetic testing laboratory. Eur J Hum Genet 18 (Suppl. 1), S1–S19.*

initiatives. The most remarkable difference between the schemes organized in the United States and Europe is the inclusion of the report in the assessment: The European schemes include the evaluation of the reports and interpretation of the result, whereas the US schemes are limited to the assessment of the correct analytical result (Dequeker, 2010).

Longitudinal research demonstrated that regular participation in EQA contributes to a more complete interpretation in laboratory test reports. It seems likely that laboratories that do not participate in EQA and whose reports have never been objectively assessed by peers and against expert opinions would have relatively low interpretation scores, comparable to those that have participated in only 1 of the 3 investigated years (Berwouts, 2012a,b).

29.7 DIAGNOSTIC VALIDATION

The validation and verification of laboratory methods and procedures before their use in clinical testing are essential for providing a safe and useful service to clinicians and patients (Mattocks et al., 2010). In addition, it is a formal requirement of accreditation standards, including ISO 17025 and ISO 15189, that (genetic) tests and instruments must be validated before diagnostic use to ensure reliable results for patients, clinicians, or referring laboratories and their quality must be maintained throughout use. In other words, the laboratory must demonstrate that its tests are fit for the intended use before application to patient samples (Berwouts et al., 2010).

The process of implementing a genetic test for diagnostic use is complex and involves many levels of assessment and validation. Key components of the process, as detailed by the ACCE framework, are analytical validation, clinical validation, clinical utility and consideration of the ethical, legal, and social implications of the test (Haddow and Palomaki, 2003). After making a decision to set up a diagnostic test, the technology to be used must be chosen and built into a suitable laboratory process.

The development stage involves assessment of both the diagnostic and technical use of the process to ensure that the measurements obtained are relevant to the diagnostic question(s) and that the analyte(s) can be unambiguously identified (i.e., that there are no confounding factors). The final stage of the laboratory process is to determine whether the performance of the test meets the required diagnostic standards in terms of accuracy. Whether this is achieved by performing analytical validation or verification depends on the existence of a suitable performance specification that details the expected accuracy of the test under given conditions (Mattocks et al., 2010; Berwouts et al., 2010).

The relatively fast emergence and the great success of new technologies, such as next-generation sequencing (NGS), hail a new era in genetic diagnostics. However, the new technologies bring challenges both at the technical level (e.g., target enrichment kits for targeted and whole-exome sequencing) and in terms of data management, as well as for the interpretation of the results. These aspects warrant a consideration of what the precise role of NGS in diagnostics will be today and tomorrow, before it even sets sail and acquires the machines and the skills. This is circular, of course, because only practice will tell us how well the tool performs. Guidelines for the evaluation and validation of NGS applications for the diagnosis of genetic disorders were written by EuroGenTest/ESHG (Matthijs et al., 2016).

29.8 QUALITY IMPROVEMENT

Internal and external audits, and follow-up of nonconformities, are essential elements to ensure the quality of laboratory results. This implies the formal approach to assess performance regularly coupled with systematic efforts to improve it. The continual process leads to the improved effectiveness of activities in a laboratory. Quality improvement can be realized by analytical test improvements, process improvement, and people-based improvement. Examples of programs are Six sigma, plan-do-check-act, and total quality management. Quality improvement involves both prospective and retrospective reviews. To achieve improvement, it is important first to measure the current status before figuring out ways to improve it. This process specifically attempts to avoid attributing blame and to create systems that prevent errors from happening.

29.9 CONCLUSIONS

Implementing a QMS is a process with different phases starting from the decision of the management to implementing a QMS, choosing an appropriate standard, delegating responsibilities, and collecting information, through formalizing SOPs, offering training, and performing validation and audits. This requires a lot of energy and time for laboratory personnel. However, all of the time invested will return in the form of increased quality for the patient and confidence in test results, as well as higher efficiency and traceability (Berwouts et al., 2010).

REFERENCES

Barton, D., Klein, C., Stacey, G.N., 2004. Certified reference materials for genetic testing. In: in Fuchs, J., Podda, M. (Eds.), Encyclopedia of Medical Genomics and Proteomics, first ed. Taylor and Francis, pp. 226–231.

Berwouts, S., Morris, M.A., Dequeker, E., 2010. Approaches to quality management and accreditation in a genetic testing laboratory. Eur. J. Hum. Genet. 18 (Suppl. 1), S1–S19.

Berwouts, S., Fanning, K., Morris, M.A., Barton, D.E., Dequeker, E., 2012a. Quality assurance practices in Europe: a survey of molecular genetic testing laboratories. Eur. J. Hum. Genet. 20 (11), 1118–1126.

Berwouts, S., Girodon, E., Schwarz, M., et al., 2012b. Improvement of interpretation in cystic fibrosis clinical laboratory reports: longitudinal analysis of external quality assessment data. Eur. J. Hum. Genet. 20, 1209–1215.

Burnett, D., Blair, C., 2001. Standards for the medical laboratory—harmonization and subsidiarity. Clin. Chim. Acta 309, 137–145.

Burnett, D., 2002. A Practical Guide to Accreditation in Laboratory Medicine.

Burnett, D., 2006. ISO 15189:2003 – quality management, evaluation and continual improvement. Clin. Chem. Lab. Med. 44, 733–739.

Canadian Standards Association: The ISO 15189:2003 Essentials – a Practical Handbook for Implementing the ISO 15189:2003 Standard for Medical Laboratories, 2004.

Cassiman, J.J., 2005. Research network: EuroGentest – a European Network of Excellence aimed at harmonizing genetic testing services. Eur. J. Hum. Genet. 13, 1103–1105.

CLSI, 2012. Quality Management for Molecular Genetic Testing. CLSI Document MM20-A. Chen B., Dequeker E.. Clinical and Laboratory Standard Institute.

Dequeker, E., Cassiman, J.J., 1998. Evaluation of CFTR gene mutation testing methods in 136 diagnostic laboratories: report of a large European external quality assessment. Eur. J. Hum. Genet. 6, 165–175.

Dequeker, E., Ramsden, S., Grody, W.W., et al., 2001. Quality control in molecular genetic testing. Nat. Rev. Genet. 2, 717–723.

Dequeker, E., 2010. Quality management systems and accreditation. In: Kristoffersson, U., Schmidtke, J., Cassiman, J.J. (Eds.), Quality Issues in Clinical Genetics Services. Springer, New York.

European Science and Technology Observatory Network. Towards Quality Assurance and Harmonisation of Genetic Testing Services in the EU, 2003.

Haddow, J.E., Palomaki, G.E., 2003. ACCE: a model process for evaluating data on emerging genetic tests. In: Khoury, M., Little, J., Burke, W. (Eds.), Human Genome Epidemiology: A Scientific Foundation for Using Genetic Information to Improve Health and Prevent Disease. Oxford University Press, New York, pp. 217–233.

Harper, J.C., Sengupta, S., Vesela, K., et al., 2010. Accreditation of the PGD laboratory. Hum. Reprod. 25, 1051–1065.

Ibarreta, D., Elles, R., Cassiman, J.J., et al., 2004. Towards quality assurance and harmonization of genetic testing services in the European Union. Nature Biotechnol. 22, 1230–1235.

International Organization for Standardization: ISO 15189 Medical Laboratories – Particular Requirements for Quality and Competence, 2012.

Kenny, D., 2001. ISO and CEN documents on quality in medical laboratories. Clin. Chim. Acta 309, 121–125.

Mattocks, C.J., Morris, M.A., Matthijs, G., Swinnen, E., Corveleyn, A., Dequeker, E., Müller, C.R., Pratt, V., Wallace, A., EuroGentest Validation Group, December 2010. A standardized framework for the validation and verification of clinical molecular genetic tests. Eur. J. Hum. Genet. 18 (12), 1276–1288.

Matthijs, G., Souche, E., Alders, M., et al., 2016. Guidelines for diagnostic next-generation sequencing. Eur. J. Hum. Genet. 24, 2–5.

McGovern, M.M., Benach, M.O., Wallenstein, S., et al., 1999. Quality assurance in molecular genetic testing laboratories. JAMA 281, 835–840.

McGovern, M.M., Benach, M., Wallenstein, S., et al., 2003. Personnel standards and quality assurance practices of biochemical genetic testing laboratories in the United States. Arch. Pathol. Lab. Med. 127, 71–76.

Muëller, C.R., 2001. Quality control in mutation analysis: the European Molecular Genetics Quality Network (EMQN). Eur. J. Pediatr. 160, 464–467.

Organisation for Economic Cooperation and Development. Quality Assurance and Proficiency Testing for Molecular Genetic Testing: Summary Results of a Survey of 18 OECD Member Countries, 2005. Organisation for Economic Co-operation and Development (OECD), Paris.

Organisation for Economic Co-operation and Development. OECD Guidelines for Quality Assurance in Molecular Genetic Testing, 2007.

Seneca, S., Morris, M.A., Patton, S., et al., 2008. Experience and outcome of 3 years of a European EQA scheme for genetic testing of the spinocerebellar ataxias. Eur. J. Hum. Genet. 16, 913–920.

Touitou, I., Rittore, C., Philibert, L., et al., 2009. An international external quality assessment for molecular diagnosis of hereditary recurrent fevers: a 3-year scheme demonstrates the need for improvement. Eur. J. Hum. Genet. 17, 890–896.

Tembuyser, L., Van Campenhout, C., Blanckaert, N., Dequeker, E.M.C., 2016. ISO 15189-accredited laboratories fulfill the JCI Hospital Accreditation Standard requirements for the use of referral laboratories: report of a consensus meeting. Accred. Qual. Assur. 1–7. (Accepted for publication).

United States Department of Health and Human Services. Notice of Intent: Genetic Testing under the Clinical Laboratory Improvement Amendments, 2011.

Westgard, J.O., 2003. Internal quality control: planning and implementation strategies. Ann. Clin. Biochem. 40, 593–611.

Westgard, J.O., 2006. Statistical quality control for quantitative measurement procedures: principles and definitions. Approv. Guidel. A3–C24.

Williams, L.O., Cole, E.C., Lubin, I.M., et al., 2003. Quality assurance in human molecular genetics testing: status and recommendations. Arch. Pathol. Lab. Med. 127, 1353–1358.

Index

'*Note*: Page numbers followed by "f" indicate figures and "t" indicate tables.'

A

Acute coronary syndromes (ACS), 242
Adverse drug effects/reactions (ADRs), 281
Allele-specific oligonucleotide (ASO) hybridization, 25–26
Amplification refractory mutation system (ARMS), 23–25, 24f

B

BioChipSet Cassette (BCSC), 66–67, 67f
Bioinformatics tools
  molecular pathway analysis, 341–350, 341f–343f
    gene regulatory networks topology information, 344–346, 345f
    gene regulatory network topology and regulatory mechanisms, 348–349
    gene regulatory network topology information, 347–348
    gene set analysis, 343–346, 344f
    integrated molecular pathway analysis system, 349–350
    MinePath web-based system, 349f
    phenotype-differentiating pathways and subpathways, 346–347
  next-generation sequencing
    bioinformatics packages, 340–341
    pipelines, 339–341
  overview, 339
Biorepository
  biohazards, 485–487, 486f
  chemical hazards, 487–490
    hazard statements, 487
    pictograms, 487, 488f
    precautionary statements, 487
    product identifier, 487
    signal words, 487
    supplier identification, 487
  electrical safety, 491
  fire safety, 491
  other areas of safety, 491
  oversight individuals involvement, 480–485
  overview, 479–480
  physical security and safety, 490–491
  regulatory and safety issues, 480, 481t–484t
  safety training/employee education, 485
B-type natriuretic peptide (BNP), 242

C

Candidate gene regions analysis, 118, 119f
Capillary electrophoresis (CE), 59–60, 200
Chromosomal translocations, 54
Circle-to-circle amplification (C2CA), 203
CLiPPER system, 347–348
Commercially available analysis platforms, DNA sequencing, 142–148
  BGI, 144
  complete genomics, 144
  illumina platform, 142
  Ion Torrent technology, 142–143
  Oxford Nanopore platform, 144–145
  pacific biosystems, 143–144
  protein nanopores, 145–146
  solid-state/graphene nanopores, 145–146
Comparative genomic hybridization (CGH), 329–330
Copy number variants (CNV), 311, 317t
Cystic fibrosis transmembrane regulator (CFTR), 233
Cystic fibrosis variant scanning, 91–92, 93f
Cytogenetics, 212
Cytogenomics
  applications
    actionable hot-spot gene panel, 275–276
    chromosomal aberrations, 275, 276f
    exome-wide mutation analysis, 276
  bioinformatics and data analysis, 277–278
  implementation, 277
  liquid biopsy, 274–275
  microscopic examination, 274
  translocation detection, 277
  tumor, 273–274

D

Data analysis, high-resolution melting curve analysis, 83–85, 84f
  background removal, 83
  curve overlay, 83
  derivative plots, 84
  difference plots, 84–85
  normalization, 83
Denaturing gradient gel electrophoresis (DGGE), 33–34, 34t
Denaturing high-performance liquid chromatography (DHPLC), 4
De Novo Discovery of Dysregulated Pathways in Human Diseases (DEGAS) methodology, 347
Designed ankyrin repeat proteins (DARPins), 220
Differential Dependency Network (DDN), 346
Digital polymerase chain reaction, 200–201, 201f
Direct to consumer (DTC) testing, 8, 334, 470–471
DNA methylation
  biology, 104–105, 104f
  clinical applications, 107–109
  disease, 106–107
    cancer, 106–107, 107f
  dynamics, 105–106
  methods, 109–126
    candidate gene regions analysis, 118, 119f
    content sample, 109–112, 110t–111t
    genome-wide analysis, 112–118
    heavymethyl and headloop polymerase chain reaction, 125f, 126
    individual CPG positions analysis, 122–123
    methylation-specific polymerase chain reaction, 123–126, 125f
    methylight, 125–126, 125f
    quantitative analysis of methylated alleles, 126
    real-time polymerase chain reaction methods, 123–126, 125f
    sequencing technologies, 118–122
  overview, 103–107, 104f
  single-cell dna methylation analysis, 126–127
DNA microarrays
  defined, 328–334
  genetic testing, 334–336
  overview, 327–328
  solid phase/two-dimensional DNA microarrays, 328–331, 329f
  suspension DNA Microarrays, 331–334, 332f–333f, 334t
DNAscan™, 66, 67f
DNA sequencing
  BioNano Genomics, 147–148
  commercially available analysis platforms, 142–148
    BGI, 144
    complete genomics, 144
    illumina platform, 142
    Ion Torrent technology, 142–143
    Oxford Nanopore platform, 144–145

DNA sequencing (*Continued*)
pacific biosystems, 143–144
protein nanopores, 145–146
solid-state/graphene nanopores, 145–146
diagnostics, chips for, 151
Genia, 146–147, 146f
overview, 141
personal genomics, 151
Roche, 146–147, 146f
techniques/systems
direct genomics single-molecule sequencing, 148–149
electron microscopy (EM), 149–150
fluorescence resonance energy transfer (FRET), 150–151
GnuBIO platform, 148
mass spectrometry (MS), 149
Qiagen platform, 148
RNA polymerase sequencing technique, 150
single DNA molecule, 149
tunneling currents, 149
DNA typing
closed accidents
in-flight airplane explosion, 403
conventional identification criteria, 398
degradation and contamination, 400–401
forensic genetics, 403–404
future perspectives, 404
human remains retrieval, 396–398, 396t, 397f
mass disasters, 401
analyzed cases description, 401–403
Bombing of Argentine Israeli Mutual Association, 1994, 402
Embassy of Israel in Buenos Aires, 401–402
murder case, 402
terrorist attack cases, 401–402
mass fatalities and diverse scenarios, 396–398, 396t, 397f
open road accidents
bus crash accident, 402
open road airplane crash, 402–403
overview, 395–396
polymorphisms, 399–400
preservation criteria, 398–399
Double-stranded DNA dyes, 81, 81f
DruGeVar database, 460, 461f

E

Electrospray ionization–mass spectrometry (ESI-MS), 204–207, 205f
Enzyme-linked immunosorbent assay (ELISA), 220–221
ESI-MS. *See* Electrospray ionization-mass spectrometry (ESI-MS)

F

Factor V Leiden genotyping, 90
FilmArray® system, 62, 62f
Fluorescence in situ hybridization (FISH), 203–204, 251, 413–414
antibody, 259
centromere-specific multicolor, 261, 262f
centromeric multiplex fluorescence in situ hybridization, 261, 262f
centromeric satellite probes, 254–255, 257f
chromosome microdissection, probe generation by, 255–256, 258f
chromosomes, 259
DNA probes, 253–254
fluorescence immunophenotyping, 259
interphase cytogenetics, 259
locus-specific probes, 256–258
morphology, 259
multicolor, 259–261, 260t
multicolor assays for subtelomeric rearrangements, 261
Primed in situ labeling (PRINS), 259
probe labeling, 254
region and locus-specific probes, 256–259
single-copy fluorescence in situ hybridization probes, 259
in situ hybridization, 259
subtelomeric probes and telomere repeat probes, 259
targets, 252–253
tumor probes, 258
types, 254–259, 255f, 256t
whole-chromosome painting probes, 255
Fluorescence resonance energy transfer (FRET), 43–44
Fluorescent SSCP (F-SSCP), 27–28
Forensic genetics, 54
Forensic science, DNA
capillary electrophoresis and data interpretation, 384–386
interpretation of results, 384–386, 385f
development and advantages, 372–373, 373f
extraction methodologies, 379–382
chelation extraction, 379, 380f
differential extraction, 381–382, 381f
organic (phenol-chloroform) extraction, 379
solid phase extraction, 381
genetic markers, 373–379
mitochondrial DNA markers, 377–378
short tandem repeat (STR) markers, 373–376, 374f, 376f
single nucleotide polymorphism markers, 378–379
Y chromosome markers, 376–377
human identity testing, 371–372, 372t
next generation, 388–391
next-generation sequencing, 390–391
rapid DNA profiling, 388–390
overview, 371–373
quantitation, 382–383, 383f
commercial quantitative polymerase chain reaction kits, 382–383
real-time quantitative polymerase chain reaction, 382
statistical calculations
counting method, 387–388
inclusion/exclusion combined probability, 387
likelihood ratio, 387
random match probability, 386–387

G

Gap-fill probes, 174–175
high-throughput targeted sequencing using, 182–184
infectious disease diagnostics, 182
Gastrointestinal (GI) pathogens, 68–71
Gene Graph Enrichment Analysis (GGEA), 348
Gene therapy, 54
Genetically modified (GM) products, 5
Genetic determinism, 471–472
Genetic epidemiology, 306–308, 307f
Genetic literacy, 469–470
Genetic scanning methods, 27–34
denaturing gradient gel electrophoresis (DGGE), 33–34, 34t
heteroduplex analysis (HDA), 30–31
single-stranded conformation polymorphism (SSCP) analysis, 27–30, 28f
temperature-gradient gel electrophoresis (TGGE), 31–33, 32f
Genetic screening methods, 23–27
allele-specific oligonucleotide (ASO) hybridization, 25–26
amplification refractory mutation system, 23–25, 24f
oligonucleotide ligation assays, 26–27
Genetic testing
awareness/access and advertising, 446–447
access to testing, 447
genetic testing, 446–447
tests, 446
challenges, 453
family communication, 452–453
individual factors, 447–450
distress, 448–449
health behaviors, 450
health beliefs, 449–450
risk perception, 448
literacy
crime scene investigation science, 472–473
defined, 469–470
genetic determinism, 471–472
health care professionals' understanding, 471
lack of public awareness, 474–475
obstacles, 471–475
public understandings, 469–471
science communication, 474–475
overview, 445–446
personal impact and professional communication, 450–452
testing negative, 451
testing positive, 450–451
unknown significance variants, 451–452
Genome Analysis Toolkit (GATK), 339–340

Genome-wide analysis, 112–118
epigenotyping arrays, 117–118
genome-wide bisulfite sequencing, 114–117, 115f
methylated DNA immunoprecipitation, 113–114
methylation reversal approach, 112
methylation-sensitive and dependent restriction enzyme-based methods, 112–113
methyl binding protein affinity chromatography, 113–114
Genomic databases
database management models, 354
future challenges, 365–366
historical overview, 353–354
locus-specific databases, 355–358
causative genomic variants, 356
comparison, 357–358
database management systems, 361–363, 362f–363f
optimal mutation detection strategy, 357
phenotypic patterns, linking genotype information with, 356–357
microattribution approach, 363–365, 365t
mutation database types, 354–355
national/ethnic mutation databases, 358–361
database management systems, 361–363, 362f–363f
genetic disease summaries module, 359–360
genetic testing, 360–361
pharmacogenomics biomarker module, 359
society, 361
overview, 353
Genomic medicine
Euro-PGx Project, 459–460
next-generation sequencing, 463
overview, 459–464
pharmacogenomics cost, 463–464
pharmacogenomics implementation, 460–462, 461f–462f
stakeholders mapping, 462–463
Genotyping, 85–88
molecular diagnostics, amplicon genotyping in, 87–88
small polymerase chain reaction products, 85–86, 85f, 86t
snapback primer genotyping, 87, 88f
unlabeled probe genotyping, 86–87, 87f
Gilbert syndrome genotyping, 90–91, 91f

H

Hazard statements, 487
Heavymethyl/headloop polymerase chain reaction, 125f, 126
Heteroduplex analysis (HDA), 3, 30–31
High-performance computing (HPC), 340
2D high-performance liquid chromatography (2D-HPLC), 237–239
High-resolution melting curve analysis (HRMA), 201–203, 202t
clinical diagnostics examples
cystic fibrosis variant scanning, 91–92, 93f
Factor V Leiden genotyping, 90
Gilbert syndrome genotyping, 90–91, 91f
mitochondrial genome variant scanning, 92
components, 80–85
data analysis, 83–85, 84f
double-stranded DNA dyes, 81, 81f
instrumentation, 81–83, 82f
copy number analysis, 94
defined, 79–85
DNA methylation detection, 201–203, 202f
genotyping, 85–88
molecular diagnostics, amplicon genotyping in, 87–88
small polymerase chain reaction products, 85–86, 85f, 86t
snapback primer genotyping, 87, 88f
unlabeled probe genotyping, 86–87, 87f
history, 79–80
methylation analysis, 92–94
padlock/selector probes, 203
prediction and assay design tools, 96–97, 96f
rare allele detection, 94
sequence matching, 94–96
genetic mapping and editing, 96
human leukocyte antigen matching, 95
repeated sequences, 95–96
variant scanning, 88–90
molecular diagnostics applications, 89–90
Human Genome Variation Society (HGVS), DNA variants
authorization, 13
coding residues, 15–16, 15t
descriptions, 16–18
alleles, 17
conversions, 17
deletions, 16
duplications, 16
indels, 17
insertions, 16–17
inversions, 17
protein, 17–18
repeated sequences variability, 17
RNA, 17
substitutions, 16
translocations, 17
history, 13
Locus Reference Genomic (LRG) sequence, 15
mutalyzer, 18–21, 18f–21f
numbering residues, 15
principles, 14
reference sequences, 14–15, 14t
Human identity profiling system, 65–67, 65f, 67f
Human Leukocyte Antigen (HLA), 410–411

I

Immunohistochemistry, 221
Individual CPG positions analysis
methylation-sensitive single-nucleotide primer extension assays, 122–123
restriction enzyme-based analysis, 122
Infectious disease diagnostics system, 61–65, 61t, 62f
Infectious pathogen diagnostics
central nervous system (CNS) pathogens, 73
gastrointestinal pathogens, 68–72
respiratory pathogens, 67–68, 69t–70t
sexually transmitted pathogens, 72–73
Inner cell mass (ICM), 409
Integrated polymerase chain reaction technologies
clinical applications
infectious pathogen diagnostics. *See* Infectious pathogen diagnostics
commercial sample-to-answer assay systems, 61–67
human identity profiling system, 65–67, 65f, 67f
infectious disease diagnostics system, 61–65, 61t, 62f
continuing evolution, 75–76
defined, 60–61
forensic applications
human identity testing, 73–75
standard workflow, 59–60
Internal quality control (IQC), 495–496, 498
International Standards for Cytogenomic Arrays (ICSA), 331

K

Kyoto Encyclopedia of Genes and Genomes (KEGG), 344

L

Linkage disequilibrium (LD), 283
Lita™, 64
Locked nucleic acid (LNA), 47–49, 333–334
Loop-mediated isothermal amplification (LAMP), 210
Low- and medium-throughput variant detection methods
genetic scanning methods, 27–34
denaturing gradient gel electrophoresis (DGGE), 33–34, 34t
heteroduplex analysis (HDA), 30–31
single-stranded conformation polymorphism (SSCP) analysis, 27–30, 28f
temperature-gradient gel electrophoresis (TGGE), 31–33, 32f
genetic screening methods, 23–27
allele-specific oligonucleotide (ASO) hybridization, 25–26
amplification refractory mutation system, 23–25, 24f
oligonucleotide ligation assays, 26–27
overview, 23

M

Matrix-assisted laser deposition and ionization-mass spectrometry (MALDI-MS), 206f, 207–208
6-Mercaptopurine (6-MP), 282
Methylation analysis, 92–94
Methylation-specific polymerase chain reaction, 123–126, 125f
Microfluidics
  cell sorting, 212
  cytogenetics, 203–204
  DNA amplification and analysis, 198–201, 198t, 199f
    capillary electrophoresis (CE), 200
    digital polymerase chain reaction, 200–201, 201f
    mutation detection, 199–200
    purification, 200
    Sanger sequencing, 200
  fluorescence in situ hybridization (FISH) analysis, 203–204
  future, 212–213
  highresolution melting analysis, 201–203, 202t
    DNA methylation detection, 201–203, 202f
    padlock/selector probes, 203
  overview, 197–198
  protein detection and analysis, 204–208
    electrospray ionization–mass spectrometry (ESI-MS) integrated microfluidic platforms, 204–207, 205f
    matrix-assisted laser deposition and ionization–mass spectrometry (MALDI-MS), 206f, 207–208
    popular microfluidic detection platforms, 208
  sample preparation, 208–212, 209t
    cytogenetics, 212
    isothermal amplification, 210, 211f
    polymerase chain reaction, 209–210
    sequencing, 210–212, 211f
Minimal residual disease (MRD), 53
Minimum Information About a Microarray Experiment (MIAME), 330–331
Minor groove binding probes (MGB), 44
Mitochondrial genome variant scanning, 92
Molecular beacons, 46
Molecular cytogenetics
  array-based techniques, 266–267, 266f
  comparative genomic hybridization, 263–266, 265f
  conventional to, 250–251
  detection and documentation, 254
  fluorescence in situ hybridization (FISH), 251
    antibody, 259
    centromere-specific multicolor, 261, 262f
    centromeric multiplex fluorescence in situ hybridization, 261, 262f
    centromeric satellite probes, 254–255, 257f
    chromosome microdissection, probe generation by, 255–256, 258f
    chromosomes, 259
    DNA probes, 253–254
    fluorescence immunophenotyping, 259
    interphase cytogenetics, 259
    locus-specific probes, 256–258
    morphology, 259
    multicolor, 259–261, 260t
    multicolor assays for subtelomeric rearrangements, 261
    Primed in situ labeling (PRINS), 259
    probe labeling, 254
    region and locus-specific probes, 256–259
    single-copy fluorescence in situ hybridization probes, 259
    in situ hybridization, 259
    subtelomeric probes and telomere repeat probes, 259
    targets, 252–253
    tumor probes, 258
    types, 254–259, 255f, 256t
    whole-chromosome painting probes, 255
  future perspectives, 267–268
  hybridization, 254
  multicolor chromosome banding techniques, 262–263, 264f
  multicolor whole-metaphase scanning techniques, 261–262
  overview, 249–250, 250f
  posthybridization wash, 254
  technical elements and materials, 251–254, 253f
  whole-genome scanning, 263–266, 265f
Molecular diagnostics
  commercializing, 6–7
  future perspectives, 6–8
  molecular biology discoveries, 1, 2t
  overview, 1
  personal genomics, 8
  personalized medicine, 7–8
  post-genomic era, 4–5
  post-polymerase chain reaction revolution, 3–4
    electrophoretic-based techniques, 3
    enzymatic-based methods, 3
    solid phase-based techniques, 3–4
Molecular inversion probes
  defined, 174–175
  genotyping application, 176–177
    massively multiplex targeted genotyping, 176–177
  infectious disease diagnostics, 182
Multiple displacement amplification (MDA), 412–413
Multiple-reaction monitoring (MRM), 234–235, 240
Mycobacterium tuberculosis, 68

N

Nature of scientific knowledge (NOSK), 473
Next-generation sequencing (NGS), 335
Noninvasive prenatal testing (NIPT)
  biological and technical factors, 432–434
    chromosomal value, 433
    chromosome guanine cytosine base content, 433
    fetal fraction, 432–433
    maternal copy number variations, 434
    mosaicism, 433, 434f
  cell-free fetal DNA, 427, 427f
  clinical setting, 436–437
  clinical trials, 434–436, 435t
    aneuploidies, 435–436
    twin pregnancies, 435–436
  ethical issues, 437–438
  false positive and false negative results, 436
  fetal-maternal differentially methylated regions, 426
  future applications, 438
  historical background, 424–427
  maternal circulation
    cell-free fetal DNA, 425–426, 425f
    fetal cells in, 424–425
    free-fetal RNA, 426
  methodologies, 427–432
    microarray-based targeted methodology, 432, 432f
    single nucleotide polymorphism-based targeted sequencing, 429–431, 430f
    targeted capture enrichment methodology, 431–432, 431f
    whole-genome sequencing, 428–429, 429f
  next-generation sequencing, 426–427
  overview, 423
  prenatal screening and diagnosis practices, 423–424
    diagnosis, 424
    hazards, 424
    screening, 423–424
Non–small cell lung cancer (NSCLC), 242
Nutrigenomics
  bioinformatics, 320–321
  data management, 320–321
  experimental models, 312–315
    animal models, 312–314
    challenging homeostasis, 315
    high-throughput assays, 312
    human clinical trials, 314–315
    interleukin 10 gene deficient ($IL10^{-/-}$), 313
    multidrug resistance gene deficient ($MDR1A^{-/-}$) mouse, 313–314
    subject preselection/stratification, 315
  genetic variation, 306–311
    copy number variants (CNV), 311, 317t
    genetic epidemiology, 306–308, 307f
    single nucleotide polymorphisms, 308–311
  integrating complex data sets, 320–321
  nutritional epidemiology, 311–312
  overview, 305–306

phenotype, 315–320, 317f
biomarkers, 320
epigenomics, 318–319
metabolomics, 319
microbiota analysis, 319–320
proteomics, 319
transcriptomics, 316–318
statistics, 320–321

O

Oligonucleotide ligation assays, 26–27
Overrepresentation analysis (ORA) approaches, 342

P

Padlock probes
automation and miniaturization, 189–190
biosensor approaches based, 177–179
defined, 173–176
DNA detection in microbes, 189
DNA nanowires, 179
genotyping application, 176–177
hyperbranched rolling circle amplification, 176
genotyping in situ, 185
high-multiplexing assays, in situ sequencing for, 188–189
homogenous amplified single-molecule detection, 177–178, 178f
infectious disease diagnostics, 179–182, 181f
magnetic biosensor approaches, 179
nuclear DNA detection, 185–188
RNA in situ detection, 188
rolling circle–amplified padlock probes, 177–179
single-molecule detection, 185, 186f–187f
in situ nucleic acid detection, 184–189
Paper-based detection platforms, 335
Peptide nucleic acids (PNAs), 47–49
Pharmacogenomics
analytical methods, 283
clinical settings, 283–287
abacavir, 286–287
CYP2C9/VKORC1, 283–284
CYP2D6, 285
*HLA*B5710*, 286–287
irinotecan, 285–286
6-mercaptopurine, 284–285
tamoxifen, 285
TPMT, 284–285
UGT1A1, 285–286
warfarin, 283–284
complex phenotypes, 288–289
drug development, 290, 291t
ethical implications, 294–295
pharmacogenomics legislation, 294
secondary information, 294–295
future perspectives, 296
history, 282–283
new trends, 292–294
overview, 281
pharmacogenetics *vs.*, 281–282, 282f
pharmacometabolomics-aided pharmacogenomics, 292–293
population differences, 287–288
public health, 295–296
regulatory agencies, 289–290
useful resources in, 292, 293t
whole genome sequencing, 293–294
Pictograms, 487, 488f
Polymerase chain reaction (PCR). *See* Quantitative polymerase chain reaction (qPCR)
Preimplantation genetic diagnosis (PGD)
accuracy, 418
assisted reproductive technology and biopsy, 408–410
blastocyst biopsy, 409–410, 409f–410f
cleavage-stage biopsy, 408–409
polar bodies (PBs), 408
in vitro fertilization, 408
chromosomal aberrations
indications, 413
clinical outcome, 416–417, 417f
emerging technologies, 415–416
future perspectives, 418
monogenic disorders
comparative genomic hybridization and array-comparative genomic hybridization-based analysis, 414–415
fluorescence in situ hybridization-based analysis, 413–414
indications, 410–411
polymerase chain reaction-based analysis, 414
targeted polymerase chain reaction-based amplification, 411–412, 412f
whole-genome amplification, 412–413
overview, 407–408
quality control, 418
Preimplantation genetic screening (PGS), 407
Primed in situ labeling (PRINS), 259
Probes, 86–87
Product identifier, 487
Protein
extensions, 18
frame shifts, 18
nonsense variants, 18
Protein diagnostics
current affinity-based protein detection assays, 220–221, 221f
future perspectives, 227–228
overview, 219–220
proteome binding, 220
proximity-dependent assays
detect proteins in situ, 225–227, 226f
detect proteins in solution, 222–225, 223f
proximity-dependent nucleic acid-based assays, 221–227
proximity extension assay (PEA), 223–225, 224f
Proteomics
bioinformatics, 240, 240f
challenges, 243
clinical impact, 233–234, 234f
discovery and verification proteomics, 240, 241t
future advances, 243–244
mass spectrometry–based proteomics, 234–239
absolute quantitation strategies, 239
basic workflow, 234–235, 235f
instrumentation, 236, 237f
quantitative mass spectrometry, 236–239
sample preparation, 235–236
overview, 233
protein-based diagnostics assays, 240–243

Q

Quality assurance, genetic laboratories
accreditation and certification, 495
diagnostic validation, 499
document control, 497–498, 498f
international standards, 494–495
overview, 493–494
quality assessment, 498–499
quality control, 498
quality improvement, 499
quality management system, 495–498, 496f–497f
elements, 495–498, 497f
organization, 496–497
Quantitative analysis of methylated alleles, 126
Quantitative mass spectrometry
gel electrophoresis, 237
label-free quantification, 239
labeling strategies, 239
Quantitative polymerase chain reaction (qPCR)
amplification reaction, 41, 42f
history, 41–43
multiplex quantitative polymerase chain reaction, 52–53
obtained data, 49–50, 50f
optimizing criteria, 54–56
principles, 41, 42f, 43–49
DNA intercalating dyes, 44–46, 45f
hybridization probes, 46, 47f
hydrolysis probes, 43–44, 44f
modified nucleic acids, 47–49, 48f
primer probes, 47, 48f
quantified data, 50–52
comparative threshold cycle method, 52
Pfaffl method, 52
standard curve method, 51, 51f
real-time thermal cyclers, 49
reverse transcriptase-quantitative polymerase chain reaction, 53–54
chromosomal translocations, 54
clinical microbiology, 53
food and industrial microbiology, 53
forensic genetics, 54
gene therapy, 54
minimal residual disease (MRD), 53
-omics techniques, 54
single nucleotide polymorphisms (SNPs), 53–54

R

RapidHIT™ 200 System, 65–66, 65f
Rare allele detection, 94
Real-time polymerase chain reaction methods, 123–126, 125f
Repeated sequences, 95–96
Restriction fragment length polymorphism (RFLP), 1, 23
Reverse transcriptase-quantitative polymerase chain reaction, 43, 53–54
  chromosomal translocations, 54
  clinical microbiology, 53
  food and industrial microbiology, 53
  forensic genetics, 54
  gene therapy, 54
  minimal residual disease (MRD), 53
  -omics techniques, 54
  single nucleotide polymorphisms (SNPs), 53–54
RNA Process Control (RPC), 63
Rolling circle amplification (RCA), 203
RT-PCR, 63–64

S

Sanger sequencing, 118, 200
Selector probes
  defined, 173–176
  high-throughput targeted sequencing using, 182–184
  multiple ligation-dependent genome amplification assay, 182, 183f
  targeted multiplex copy number variation analysis, 182
Sequence matching, 94–96
  genetic mapping and editing, 96
  human leukocyte antigen matching, 95
  repeated sequences, 95–96
Sequencing technologies
  bisulfite-treated DNA
    next-generation sequencing, 119–120
    Sanger sequencing, 118
  matrix-assisted laser desorption/ionization (MALDI), 120–122
  Pyrosequencing™, 120, 121f
Sequential window acquisition of all theoretical fragment ion spectra (SWATH), 243–244
Sexually transmitted infections (STIs), 72
Short tandem repeats (STRs), 60, 373–376, 374f, 376f
Signal words, 487
Single nucleotide polymorphisms (SNPs), 44, 53–54, 308–311
  discovery, 308–309
  nutrigenomics, 309–311, 309t–310t
Single-strand conformation polymorphism (SSCP), 3, 27–30, 28f
  denaturation, 29
  detection, 29
  DNA amplification, 29
  electrophoretic conditions, 29
Small polymerase chain reaction products, 85–86, 85f, 86t
Snapback primer genotyping, 87, 88f
Sodium dodecyl sulfate– polyacrylamide gel electrophoresis (SDS-PAGE), 220
SSCP. *See* Single-stranded conformation polymorphism (SSCP)
Standard operating procedures (SOPs), 495
Supplier identification, 487
Suspension arrays, 332

T

Target sequencing microarrays (TSMs), 463
Temperature-gradient gel electrophoresis (TGGE), 31–33, 32f
Threshold cycle (Ct) values, 49–50

U

Ultraperformance liquid chromatography (UPLC), 243
Universal Heteroduplex Generator (UHG), 30
Unlabeled probe genotyping, 86–87, 87f
Urinary tract infections (UTIs), 72

V

Viral infection, 73

W

WGS. *See* Whole-genome sequencing (WGS)
Whole-genome sequencing (WGS), 463
  advanced, 156–165
  defined, 155–156
  DNB arrays, 156–157, 158f
  genomic precision health vision
    genome compiler, 165
    perfecting/scaling personal whole-genome sequencing, 165
    in silico and in vitro genome program, 165–166, 166f
    virtual health coaching, 166–167, 167f
  health preservation, ultimate genetic test for, 164–165
  limitations, 158t
  long fragment read (LFR) technology, 158–159, 159f
  Moore's law *vs.* DNA sequencing, 157f
  overview, 155–156
  selecting optimal cancer therapy, ultimate cancer profiling for, 163–164
  sick kids, ultimate first-line diagnostic test, 159–160, 160t
  ultimate prenatal/preimplantation test, 161–163, 161f–162f

Z

Zip Nucleic Acids (ZNA™), 49
Zona pellucida (ZP), 408